UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE FÍSICA


# EASY EMBRYONIC STEM CELL

## A NEW TOOL TO MANIPULATE GENE EXPRESSION IN EMBRYOID BODIES

Mestrado Integrado em Engenharia Biomédica e Biofísica

Perfil em Engenharia Clínica e Instrumentação Médica

Felícia Margarida Correia da Fonseca Lopes

Dissertação Orientada por:

Dr. Cláudio Areias Franco

Prof. Dr. Pedro Cavaleiro Miranda

2016

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE FÍSICA


# EASY EMBRYONIC STEM CELL

## A NEW TOOL TO MANIPULATE GENE EXPRESSION IN EMBRYOID BODIES

Mestrado Integrado em Engenharia Biomédica e Biofísica

Perfil em Engenharia Clínica e Instrumentação Médica

Felícia Margarida Correia da Fonseca Lopes

Dissertação Orientada por:

Dr. Cláudio Areias Franco

Prof. Dr. Pedro Cavaleiro Miranda

2016

Para os meus pais.

# RESUMO

O sistema vascular permite a distribuição adequada de sangue e consequente fornecimento de oxigénio e nutrientes a todos os tecidos em vários organismos multicelulares. O sistema vascular é essencial para o desenvolvimento embrionário, o crescimento de tecidos e o processo de cicatrização. De especial interesse, muitas doenças que afectam o ser-humano, como por exemplo retinopatia diabética, isquemia, acidente vascular cerebral e crescimento de tumores sólidos, surgem devido a má formação e disfunção dos vasos sanguíneos.

A formação de uma rede vascular funcional requer a extensão dos vasos sanguíneos pré-existentes pelo processo de angiogénese de forma a alcançar os tecidos não vascularizados e, em simultâneo, a remodelação do plexo vascular primitivo amorfo criado por angiogénese numa rede hierarquizada de artérias, veias e capilares. Estes dois processos são fortemente regulados pela comunicação molecular entre as células dos tecidos e as células endoteliais (as células que revestem os vasos sanguíneos). De especial relevância, começa agora a ser evidente que as células endoteliais necessitam de coordenar os seus comportamentos individuais para permitir a organização de uma rede vascular de forma coerente e funcional. É, por isso, importante compreender os mecanismos moleculares de regulação do comportamento coordenando das células endoteliais durante as diferentes fases do processo de formação dos vasos sanguíneos, sendo que isso irá certamente criar novas possibilidades de tratamento médico de várias doenças vasculares e melhorar o impacto das terapias existentes na saúde humana.

Nos últimos anos temos assistido a um aumento significante tanto no desenvolvimento como no uso de tecnologias transgénicas, usadas em estudos que visam a nossa compreensão dos mecanismos biológicos, mas que têm também facilitado o desenvolvimento de uma vasta panóplia de modelos de doenças que têm grande impacto na abordagem das patologias humanas. As células estaminais embrionárias derivam de células totipotentes embrionárias e, como tal, servem como fonte putativa de vários tipos de células diferenciadas, dando origem a todos os tipos celulares do organismo. Existem duas características que distinguem as células estaminais embrionárias das restantes: a sua pluripotência e a sua capacidade de auto-renovação sob determinadas condições e, como tal, quando geneticamente modificadas *in vitro*, podem ser utilizadas no estudo de vários mecanismos biológicos em diferentes ensaios. Particularmente, as células estaminais embrionárias podem ser utilizadas em ensaios sobre angiogénese a partir da criação de corpos embrióides, que resultam da agregação *in vitro* destas células e que permitem testar o potencial das mesmas. A principal vantagem deste sistema, neste contexto, é poder ser manipulado a partir da utilização de células estaminais embrionárias distintas com diferentes modificações ou origem genética, criando quimeras que permitem a visualização da dinâmica das células endoteliais e o estudo da eficiência das diferentes fases do processo de formação dos vasos sanguíneos durante a angiogénese.

As tecnologias correntemente desenvolvidas para inserir um gene numa célula viva estão limitadas pela natureza aleatória desta inserção no genoma da mesma. O novo gene, ao ser posicionado arbitrariamente, pode inactivar ou perturbar o funcionamento de genes endógenos, causando efeitos indesejados na célula. Adicionalmente, estas tecnologias não permitem a reprodutibilidade do processo uma vez que não há garantia de que a nova sequência seja inserida na mesma posição do genoma em duas células diferentes. Este problema pode ser contornado utilizando tecnologias que permitem recombinação homóloga, na inserção de transgenes em locais específicos ou em modificações *in situ* de genes existentes, mas que são muito morosas pela necessidade de alcançar e seleccionar os clones positivos.

De forma ultrapassar as dificuldades supramencionadas, pretendemos aumentar a eficiência e velocidade do processo de inserção de genes em células a partir do desenvolvimento de um método de fácil utilização para a manipulação de células estaminais embrionárias para o estudo em corpos embrióides. Para tal, pretendemos recorrer à actividade de um tipo de enzimas específico que reconhece curtas sequências de ADN que medeia a recombinação entre estes elementos, resultando em excisão, integração, inversão e substituição dos fragmentos de ADN, as chamadas *site-specific* recombinases (de sequência específica). Dentro desta família de enzimas, existe uma, a integrase PhiC31, que por recombinação homóloga e através de locais específicos de fixação, consegue integrar um fragmento de ADN de qualquer dimensão, de forma específica e irreversível. Com base neste mecanismo, desenvolvemos dois vectores, cada um deles com os locais de fixação específicos, sendo que um deles possui o genoma alvo e o segundo, o gene que pretendemos introduzir na célula. Através da expressão endógena de integrase PhiC31, a sequência de interesse será introduzida no genoma da célula de forma específica e eficiente, prevenindo a disrupção de genes endógenos e oferecendo perfeita reprodutibilidade.

A principal vantagem desta ferramenta para o nosso laboratório prende-se com o facto de facilitar a manipulação de células estaminais embrionárias, que permitem a criação de corpos embrióides quiméricos para a visualização da dinâmica das células endoteliais e a eficiência de migração durante o processo angiogénico.

Até ao momento a nossa estratégia revelou ser promissora, tendo a expressão de integrase PhiC31 sido bem-sucedida em células cancerígenas em cultura mas, faltando ainda concluir a validação do vector de expressão, que revelou deformidades. A próxima fase será extrapolar a técnica para células estaminais embrionárias para que seja aprovada e posteriormente aplicada a diferentes ensaios em outras áreas da biologia.

**Palavras-Chave:** Vasos Sanguíneos, Rede Vascular, Angiogénese, Células Endoteliais, Células Estaminais Embrionárias, Corpos Embrióides, integrase PhiC31.

# ABSTRACT

The established transgenesis methods used in embryonic stem cells allow efficient genomic integration of transgenes. However, the integration of multiple copies or transgenes at random genomic locations complicates comparative transgene analysis and makes long-term transgene stability unpredictable with variable expression. Targeted, site-directed transgene integration into pre-determined genomic loci can circumvent these issues, enabling direct comparison of different transgenes, thereby improving time and cost efficiency. The PhiC31 integrase catalyzes the unidirectional recombination reaction between heterotypic attP and attB sites and is an efficient platform for site-directed transgenesis.

Here, we implemented an efficient PhiC31-based site-specific transgenesis system for embryonic stem cells to be used in embryoid bodies' studies. It allows the recombination of attB-containing transgene vectors into single genomic attP landing sites being this enzyme endogenously expressed by the cell, avoiding PhiC31 integrase encoding mRNA injection. Our strategy revealed a great potential as, cancer cells in culture were able to express PhiC31 integrase but, the validation of the expression vector is still ongoing due to deformities in the plasmid. The next step will be validating the technique in embryonic stem cells.

The major advantage of this system for our laboratory is that it can easily be manipulated by using different transgenic embryonic stem cells, creating chimeras to visualize the dynamics of endothelial cells and study the efficiency of sprouting and migration during angiogenesis.

This tool can also be applied to many other assays in transversal fields of biology. We hope that our results help researchers to optimize genome editing in their specific systems.

**Keywords:** Blood Vessels, Vascular Network, Angiogenesis, Endothelial Cells, Embryonic Stem Cells, Embryoid Bodies, PhiC31 integrase.

# ACKNOWLEDGEMENTS

There are a number of people to whom I am deeply grateful for guidance, advice, and support.

First, I would like to express my gratitude to my external supervisor, Cláudio Franco, for teaching me how to do science, for leading by example, and for pushing me to always become better. Cláudio is an exceptionally creative thinker and has continuously supported me in all my endeavors. Under his mentorship, inspired by his infectious enthusiasm, I have learned and grown more than I ever imagined.

I would also like to thank my internal supervisor, Pedro Cavaleiro Miranda, for dealing with all the paperwork implied in the realization of this thesis and for his valuable input in the written work.

Moreover, thanks to all the professors from Instituto de Biofísica e Engenharia Biomédica whom, during my academic life, were always very attentive and helpful in many different manners.

I am very grateful to all of the past and present members of the CFranco Lab, and our neighbors in the EGomes Lab, for the invaluable input, guidance and assistance, patience and friendship during my internship at Instituto de Medicina Molecular.

Importantly, I want to thank all my great friends and colleagues for all the joyful times that we spent together, for our travels and adventures, for the company during the countless nights spent studying and for being there when I most needed. I must also thank my sister for all the great moments, complicity and for reminding me to always have fun; despite the distance that separates us, you are always the closer you can be to my heart.

Most of all, I am immensely grateful to my parents, who helped me to grow, personally and professionally, throughout my school years. Thank you for your emotional and financial support, for your friendship and guidance, for all the opportunities you have given me and, specially, for being always by my side. You are never too far away from my thoughts.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **Amp** | Ampicillin |
| **att** | Attachment |
| **CAG** | Chicken beta actin |
| **Cam** | Chloramphenicol |
| **cDNA** | Complementary DNA |
| **CIP** | Calf Intestinal Alkaline Phosphatase |
| **CMV** | Cytomegalovirus |
| **$CO_2$** | Carbon Dioxide |
| **CRISPR** | Clustered Regularly Interspaced Short Palindromic Repeats |
| **crRNA** | CRISPR RNA |
| **Ct** | Cycle Threshold |
| **DABCO** | 1,4-Diazabicyclo[2.2.2]Octane |
| **DAPI** | 4',6-Diamidino-2-Phenylindole |
| **dCas9** | Nuclease-deficient Cas9 |
| **$dH_2O$** | Distilled Water |
| **Dll4** | Delta-like Ligand 4 |
| **DMEM** | Dulbecco's Modified Eagle Medium |
| **DNA** | Deoxyribonucleic Acid |
| **dNTPs** | Deoxyribonucleoside Triphosphates |
| **E. coli** | Escherichia coli |
| **EB** | Embryoid Body |
| **EDTA** | Ethylenediaminetetraacetic Acid |
| **ES** | Embryonic Stem |
| **EtOH** | Ethanol |
| **G418** | Geneticin |
| **GAPDH** | Glyceraldehyde-3-Phosphate Dehydrogenase |
| **GOI** | Gene of Interest |
| **$H_2O$** | Water |
| **Kan** | Kanamycin |
| **LB** | Luria Broth |
| **mRNA** | Messenger RNA |
| **ON** | Overnight |
| **PBS** | Phosphate Buffer Solution |
| **PCR** | Polymerase Chain Reaction |
| **PFA** | Paraformaldehyde |
| **RNA** | Ribonucleic Acid |
| **RNAi** | RNA Interference |
| **RT** | Room Temperature |
| **sgRNA** | Single Guide RNA |

| | |
|---|---|
| SSRs | Site-specific Recombinases |
| TALENs | Transcription-activator Like Effector Nucleases |
| TBE | Tris/Borate/EDTA |
| Tm | Melting Temperature |
| tracrRNA | Trans-activating crRNA |
| Tris | Tris(hydroxylmethyl)aminomethane |
| UTR | Untranslated Regions |
| UV | Ultra-violet |
| VEGF | Vascular Endothelial Growth Factor |
| VEGFR | Vascular Endothelial Growth Factor Receptor |
| WT | Wild Type |
| ZFNs | Zinc-finger Nucleases |

# LIST OF UNITS

| | |
|---|---|
| % | Percent |
| °C | Degree Celsius |
| µg/µL | Microgram per Microliter |
| µL | Microliter |
| µM | Micromolar |
| bp | Base Pair |
| h | Hour |
| kb | Kilobase Pair |
| L | Liter |
| M | Molar |
| min | Minute |
| mL | Milliliter |
| ng/µL | Nanogram per Microliter |
| rpm | Rotations per Minute |
| sec | Second |
| V | Volt |
| w/v % | Weight/Volume Percentage |

# 1. INTRODUCTION

Blood vessels have being studied for many centuries and even though we now know some of the molecular mechanism controlling the aspects that make the vascular system a very complex, stereotyped and hierarchical network, there are still a lot of unanswered questions behind the blood vessel formation. Importantly, the formation and maintenance of a functional vascular network is fundamental for growth and development of multicellular organisms as an incorrect vascular assembly is linked to several pathological states such as solid tumors, cerebral cavernous malformations, hemangiomas, retinopathies and arteriovenous malformations.

The first vessels in the developing embryo form through vasculogenesis, after which angiogenesis, the physiological process through which new blood vessels form from pre-existing ones, is responsible for most, if not all, blood vessel growth during development and in disease. During sprouting angiogenesis, there are two distinct types of cells, the tip and stalk cells, whose phenotype is known to be regulated by a VEGF/Notch signaling.

In the past years, we have witnessed a significant increase in both the development and use of transgenic technologies that have been used to aid our fundamental understanding of biologic mechanisms, but that have also facilitated the development of a range of disease models that have a huge impact upon our approach to human disease. Embryonic stem cells are derived from totipotent cells of the embryo and therefore, can serve as a putative source of numerous types of differentiated cells, thus generating every cell type in the body. Two distinctive properties distinguish embryonic stem cells, their pluripotency and their capacity for self-renewal under defined conditions. Embryonic stem cells can be genetically modified *in vitro* and thus be used to study many different mechanisms in different assays, particularly, they are used in sprouting assays by creating embryoid bodies. Embryoid bodies are the result of the *in vitro* aggregation of embryonic stem cells and are utilized to test the differential potential of these cells. The major advantage of this system for our laboratory is that it can be manipulated by using different embryonic stem cells with different modifications or distinct genetic background, creating a chimera (a single organism composed of genetically distinct cells) to visualize the dynamics of endothelial cells and study the efficiency of sprouting and migration during angiogenesis.

The current technologies used to insert a gene into a living cell are limited by the random nature of the insertion of the new sequence into the genome. The new gene is positioned arbitrarily, and may inactivate or disturb the functioning of other genes or even cause severe unwanted effects. Furthermore, these technologies offer no degree of reproducibility, as there is no guarantee that the new sequence will be inserted at the same place in two different cells. The precise genetic modification can be achieved in embryonic stem cells by taking advantage of homologous recombination to target single-copy transgenes to specific sites or to modify existing

genes *in situ*, but this technique is very time-consuming because it is necessary to target and select the positive clones.

Site-specific recombinases recognize short DNA sequences (typically between 30 and 40bp) and mediate the recombination between these elements resulting in excision, integration, inversion, or exchange of DNA fragments. The PhiC31 integrase is a sequence-specific recombinase that by homologous recombination and using attachment (att) sites, can integrate a plasmid of any size in an irreversible way. Taking advantage of these properties and in order to overcome the difficulties inherent to the existing methods for cell genome manipulation, we developed a way to easily manipulate embryonic stem cells to test embryoid bodies, increasing the efficiency and speed of the clone selection process.

We created two vectors, one with the target genome and the other with the protein that we want to express. Each of the vectors has a PhiC31 attachment site and, through the endogenous expression of PhiC31 integrase, induced by the first vector, the expression plasmid is inserted in the genome of the cell, in a specific and efficient way. The major advantages of this system is that it enables a specific area of the DNA to be modified, thereby increasing the precision of the correction or insertion, preventing any cell toxicity and offering perfect reproducibility. At the moment, our strategy revealed to have potential as the insertion of the targeting vector in U2OS cells in culture worked well and cells are expressing PhiC31 integrase but, we were still not able to create an expression vector able to confirm the efficiency of our technique. The next step is to test the constructs in embryonic stem cells to validate it so it can, posteriorly, be applied to different systems of other fields of biology.

This study took place at the Vascular Morphogenesis Laboratory, in Instituto de Medicina Molecular, Lisbon, and was performed under the supervision of Dr. Cláudio Areias Franco, group leader of the laboratory that is mainly focused on the understanding of the molecular mechanisms regulating coordinated endothelial cell behavior during sprouting and remodeling phases of the angiogenic process. Improving the knowledge on the molecular regulation of vascular morphogenesis will certainly create new possibilities for medical prevention and treatment of various human conditions.

# 2. BACKGROUND

The correct development of blood vessels is essential for all aspects of tissue growth and physiology in vertebrates as they supply oxygen and nutrients to the organism. In the adult, vessels are normally quiescent; however, structural or functional vessel abnormalities may occur. An inadequate vessel maintenance or growth causes ischemia in diseases such as myocardial infarction, stroke, and neurodegenerative or obesity-associated disorders, whereas excessive vascular growth or abnormal remodeling promotes many ailments including cancer, inflammatory disorders, and eye disease [1][2]. The formation of an elaborate hierarchically branched vascular network, through either vasculogenesis or angiogenesis, relies on a series of coordinated morphogenic events that involves complex and highly dynamic interactions between endothelial cells and their environment.

In the embryo, new vessels form de novo via the assembly of mesoderm-derived endothelial precursors, angioblasts, that differentiate into a primitive vascular network, in a process called vasculogenesis [3]. The subsequent sprouting from the pre-existing vessels, angiogenesis, leads to the stabilization of that network remodeling it into arteries, veins and capillaries [4] thus creating the complex, hierarchical and functional vascular network and we know, allowing the correct blood supply to all tissues.

## 2.1. SPROUTING ANGIOGENESIS

In order to constantly respond to metabolic and growth demands, the vascular system is highly dynamic and endothelial cells are able to sense and respond to the environmental signals, initiating the angiogenic process. Biological signals such as hypoxia, ischemia and/or blood vessel damage, promote the upregulation of pro-angiogenic factors and, attracted by these factors, endothelial cells become motile and invasive by protruding filopodia that guide the new blood vessel in a certain direction (see Figure 1A). These so-called tip cells lead new sprouts and probe the environment for guidance cues. Following tip cells, stalk cells are able to proliferate to support sprout elongation and establish the vessel. These two cell types constitute the so-called vascular sprout, which remains connected to the original vessels while migrating into the avascular tissue. Tip cells anastomose with cells from neighboring sprouts to build vessel loops and the posterior initiation of blood flow, establishment of a basement membrane, and recruitment of mural cells stabilize the new connections (see Figure 1B and C). The iteration of the sprouting process continues until pro-angiogenic signals cease, and quiescence is re-established.

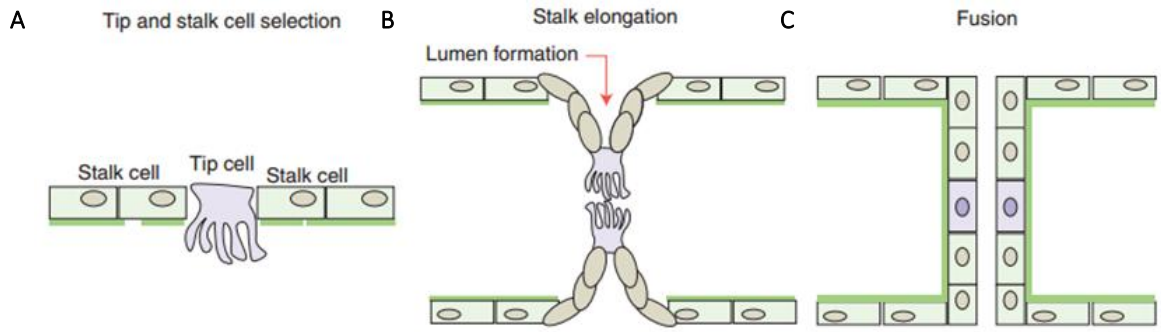



**Figure 1.** Schematic model of sprout initiation, vessel branching, and maturation. (A) Angiogenesis is activated in response to local growth factors, that (re)activates the quiescent endothelial cells differentiates them into tip and stalk cells. (B) Endothelial cells proliferate and collectively invade the tissue while remaining connected to the original vascular network and the tip cells contact other sprouts. (C) The new connection between different sprouts occurs through tip cell fusion (anastomosis) and the formation of the vascular lumen initiates blood flow. As a functional vessel loop is formed, there is a reduction of the release of endothelial growth factors, supporting the establishment of quiescence. Image adapted from [5].

### 2.1.1. VEGF AND NOTCH/DLL4 SIGNALING PATHWAYS

The specification of endothelial cells into tip and stalk cells within the activated endothelium is determined by the feedback loop between VEGF and Notch1/delta-like ligand 4 (Dll4) signaling pathways [6][7]. The resulting cell-fate specification mechanism is based on a Notch-mediated lateral inhibition process, in which a cell becomes tip cell and prevents its immediate neighbors from acquiring the same phenotype. This cell adopts this phenotype due to stochastic differences in local vascular endothelial growth factor (VEGF) concentrations, in filopodia elongation (and thus VEGF exposure) or in transcription rate that leads to small imbalances where one cell will express slightly higher Dll4 levels and, thus, dominate its neighbors by activating more Notch signaling [8][9]. The cell with more Dll4, and less Notch activity, will be selected as the tip cell. The activation of Notch inhibits VEGF receptor 2 (VEGFR2), indirectly inhibiting Dll4 expression levels, thereby reinforcing the dominance of the selected tip cell and limiting the number of tip cells induced by VEGF (see Figure 2) [10][11].

Besides downregulating VEGFR2, Notch signaling also affects VEGFR1 and VEGFR3 expression. Notch activation leads to increased levels of VEGFR1. This reduces the angiogenic sprouting response to VEGF, as both receptor variants act as a decoy for the VEGF ligand and limit VEGFR2 activation [12][13]. VEGFR3 is most strongly expressed in the leading tip cells and is downregulated by Notch signaling in the stalk cells [14]. Collectively, these molecular regulatory processes lead to increased responsiveness of the tip cell to VEGF and decreased sensitivity of stalk cells to this factor.

Recent observations challenge the idea of stable tip and stalk cell selection as their phenotype revealed to be dynamic and transient, relying on continued competition between the cells within blood vessel sprouts [15]. As a consequence, tip cells are overtaken and exchanged by neighboring stalk cells.

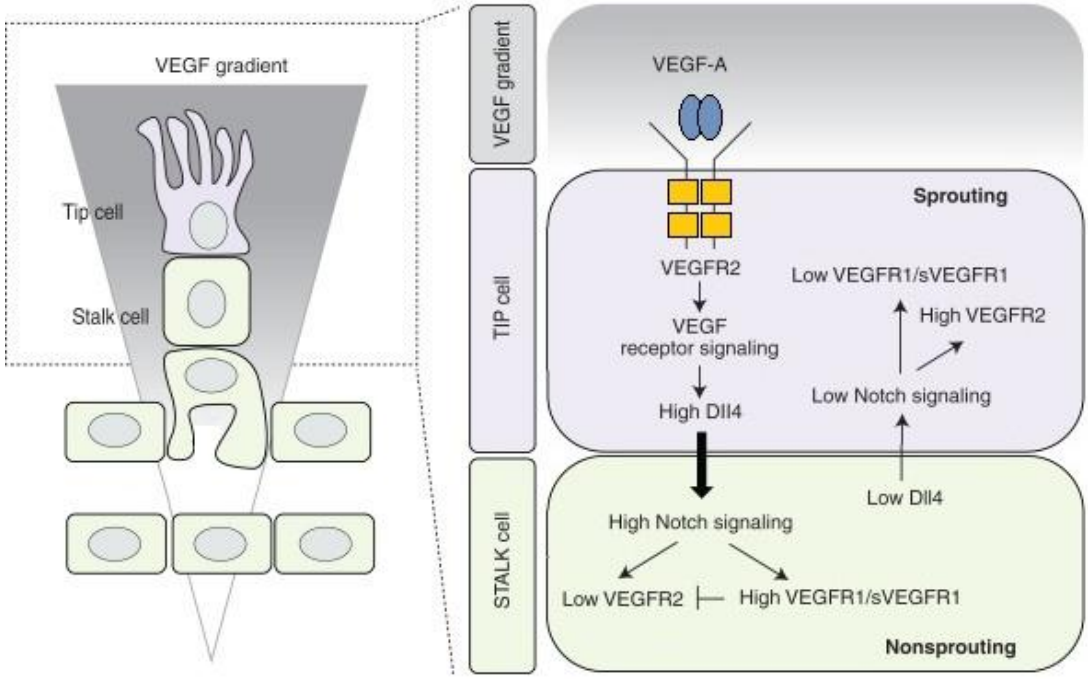



**Figure 2.** Tip to stalk cell lateral inhibition during sprouting angiogenesis. VEGF interacts with VEGFR2, expressed at the surface of the endothelial cells of the quiescent vessels. The VEGF signaling output is modulated, enhancing the binding activity and signaling of VEGF through VEGFR2. Under VEGF stimulation, Dll4 expression is up-regulated in the tip cells. In turn, Dll4 ligand activates Notch signaling in the stalk, consequently suppressing the tip cell phenotype. Notch signaling activation reduces VEGFR2 expression and increases VEGFR1/sVEGFR1 as well as the expression of different Notch target genes. In contrast, the tip cell receives low Notch signaling, allowing high expression of VEGFR2 but low VEGFR1. Image adapted from [5].

### 2.1.2. Competition between Tip and Stalk cells

Recent time-lapse imaging studies have shown the dynamic shuffling between tip and stalk cells at the leading front of growing sprouts, both *in vitro* (in embryoid bodies sprouting assays) and *in vivo* (in mice and zebrafish). The positional exchange suggests that the cells have constantly to reevaluate the VEGF/Notch signaling loop when they contact new neighboring cells, competing for the tip cell position based on their relative levels of VEGFR1 and VEGFR2. These receptors' activities mediate the expression level of the Dll4 ligand: lower expression levels of VEGFR1 or higher levels of VEGFR2 result in higher levels of Dll4 expression and, hence, an increased ability of a cell to suppress its neighboring cells from becoming tip cells [15]. This proposes that the tip cell is constantly challenged by cells within the stalk cell region of the sprout to demonstrate its dominance in terms of VEGFR levels, ensuring that the cell with the best guiding capacities leads the sprout.

### 2.1.3. Embryoid Body Sprouting Assay as a Tool to Measure Cell Competition

Embryonic stem (ES) cells are pluripotent cells derived from blastocyst-stage early mammalian embryos that are capable of differentiation into all three germ layers and, therefore, have the potential to develop any type of tissue. Because of this particular characteristic, the maintenance and differentiation of embryonic stem cells contributed greatly to significant discoveries in developmental biology. When cultured in suspension without anti-differentiation factors, embryonic stem cells spontaneously differentiate and form three-dimensional multi-cellular pluripotent aggregates called embryoid bodies (EBs) [16]. This structure facilitates multicellular interactions, in which cell-cell contact exists, and consists of a mix of cell populations with ectodermal, mesodermal, and endodermal origins. Embryoid body's aggregates recapitulate many aspects of cell differentiation during early mammalian embryogenesis offering the opportunity to mechanistically study these differentiation events of tridimensional assemblies of pluripotent cells with the great advantage that genetic manipulation of endothelial stem cells can be studied for gene mutations or knockouts that prove to be lethal during normal embryonic development *in vivo* [17]. As differentiation continues, a wide range of cell types are developed within the embryoid body's environment and, consequently, this technique has been widely utilized as a trigger during the process of embryonic stem cells differentiation *in vitro*. The embryoid bodies' potential covers different research fields and several gene knockout studies have being made using this technique. For example, embryoid bodies were used as an alternative for studying hematopoiesis [18][19], for complementary studies of cardiomyocytes differentiation [20] and for accessing endothelial cells behavior [21][22]. Currently, there are several methods with unique peculiarities used to form embryoid bodies from embryonic stem cells that allow for different objectives to be attained [23][24].

#### 2.1.3.1. Assessing Sprouting Capacity, Dynamics and Competition of Cells

The embryoid body model is valuable to easily access vessel development, by manipulation and visualization at a high resolution as the endothelial cell conventional cultures do not provide a proper microenvironment for it. Three dimensional models allow studying interactions between endothelial cells, adjacent non-endothelial cells and matrix, which are known to be essential in the regulation of vascular processes, as well as the metabolism and sprouting efficiency of these cells.

The sprouting capacity of endothelial cells during sprouting angiogenesis can be visualized by using embryoid bodies by designing a mosaic sprouting assays using embryonic stem cells. These cells, as said before, are able to form embryoid bodies that, when cultured in a solidified collagen matrix and treated with VEGF, exhibit robust differentiation into the endothelial cell lineage and an angiogenic behavior that resembles the major steps of sprouting angiogenesis *in vivo* [25][26].

Previous work described that chimeric cultures of embryonic stem cell, originated from 1:1 mixtures of two wild-type embryonic stem cell lines, revealed mosaic vascular sprouts. Deeper analysis of the genotypic origin of

the leading tip cells have shown an equal participation of each cell population to the leading tip cells, proving that wild-type cells of different genetic background have equal potential to acquire the tip cell phenotype (see Figure 3A and B)[15].

To investigate the interactions between cells in the selection process and the role of VEGFR levels in it, the same group studied different combinations of endothelial cells with half the amount of VEGFR1 or VEGFR2 competing with wild-type neighboring cells for the leading position [15]. It was shown that endothelial cells derived from embryonic stem cells heterozygous for the Vegfr2 allele (Vegfr2$^{+/egfp}$, that exhibited approximately half the VEGFR2 levels, comparing to wild-type cells), when mixed with wild-type cells in a 1:1 ratio in mosaic cultures, contributed to only around 13% of the tip cells, showing that reduced Vegfr2 levels selectively impair the ability of cells to acquire the tip phenotype. When varying Vegfr1 levels using chimeric cultures of wildtype and Vegfr1$^{+/lacz}$ embryonic stem cells (that exhibited approximately half the VEGFR1 levels in comparison to wild-type cells), 70% of the tip cells were derived from the Vegfr1 heterozygote population. This demonstrates that a cell with lower Vegfr1 expression has a higher probability of acquiring the leading position, whereas a cell with lower Vegfr2 expression has reduced ability to take the lead, when competing with a wild-type neighboring cell, suggesting that the balance of VEGFR2 and VEGFR1 expression in individual endothelial cells affects their potential to become tip cells during sprouting angiogenesis (see Figure 3).

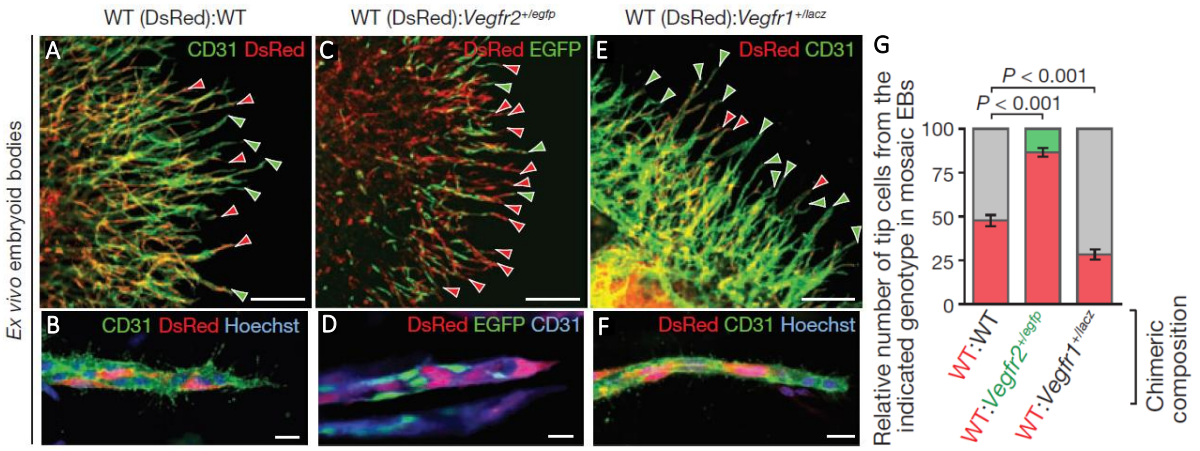



**Figure 3.** VEGFR levels in mosaic-tip-cell selection in vitro. (A, B) Sprouting vasculature (green) of a chimeric embryoid body composed of two different wild-type strains, immunolabelled with antibodies specific to CD31. One strain has ubiquitous expression of the fluorophore DsRed-MST (tip cells indicated by red arrowheads) and the other one is a non-fluorescent embryonic stem cell line (R1) (tip cells indicated by green arrowheads). (C, D) Chimeric embryoid body of wild-type (red) and Vegfr2 heterozygote cells (Vegfr2+/egfp, green) mixed in a 1:1 ratio. Wild-type tip cells are indicated by red arrowheads, Vegfr2+/egfp tip cells are indicated by green arrowheads. (E, F) Chimeric embryoid body of wild-type cells (red; tip cells are indicated by red arrowheads) and Vegfr1 heterozygote cells (tip cells are indicated by green arrrowheads), mixed in a 1:1 ratio. (G) Quantification of tip cells with the indicated genotype in chimeric embryoid bodies. Number of counted tip cells per condition: 630–708 (n embryoid bodies per condition = 6, WT:WT; 11, WT:vegfr2+/egfp; 9, WT:vegfr1+/lacz). Values represent means ± s.e.m. Scale bars: a, c, e, 200µm; b, d, f, 20µm. Image adapted from [15].

Despite embryonic stem cells and embryoid bodies assays' great potential and usefulness for this kind of studies, there is a limited number of genetically modified cell lines available. This results in a gap where knock-downs of genes, expression of fluorescent reporters and overexpression of specific isoforms, including fluorescently-tagged proteins, point mutations forms and dominant-negative/active isoforms are lacking.

## 2.2. How to Manipulate Stem Cells

Embryonic stem cells have one great advantage over other cell types that is their accessibility to genetic manipulation. These cells are able to remain pluripotent after genetic modifications and so, they are widely used and there is a range of effective techniques that have been established for gene delivery and manipulation of embryonic stem cells. These methodologies include electroporation, transfection and infection protocols, and also different approaches for inserting, deleting, or changing the expression of genes.

### 2.2.1. Genetic Engineering

The strategies used in genetic engineering usually require the permanent modification of the target genome. For this to happen, the insertion of a construct into the genome of the cell needs to be stable but the prevailing approaches for this over the last years have been based in methods that lead to a random integration. Randomly integrating technologies enable users to create stable systems leading to lasting expression which results in a random insertion of selected DNA fragments into the host genome without the use of DNA homology. Besides providing a valuable tool for long-term expression in embryonic stem cells, random insertion leads to inconsistent integration sites and unpredictable expression patterns. In addition, the locus of insertion can result in partial or complete silencing of endogenous genes of embryonic stem cells, which can result in insertional mutagenesis with subsequent genome instability or toxicity [27][28].

Site-specific modification is an ideal method to avoid variable expression patterns and copy number variation as a result of random integration. These tools enable the targeting of specific sites within the chromosome where silencing is minimized. Gene targeting using traditional methods via homologous recombination has been extensively used to specifically alter genes. This method involves the introduction of a targeting construct homologous to the target gene sequence of the embryonic stem cell [29]. Despite offering, in principle, great specificity to the integration process, homologous recombination has an extremely low efficiency and is very laborious. With this in mind, new strategies have been developed to achieve integration at predetermined target sites with high efficiency, based on site-specific recombinases like Cre, FLP and PhiC31 integrase.

#### 2.2.1.1. Site-Specific Recombinases

Site-specific recombinases (SSRs) are highly specialized enzymes that perform rearrangements of DNA fragments by recognizing and binding to short DNA sequences, sites. Recombination sites are typically between

30 and 200 nucleotides to which the recombinase binds, flanking a central crossover sequence at which the recombination takes place [30][31][32].

FLP and Cre recombinases belong to the integrase family of recombinases (also termed "λ integrase" family) and do not need any accessory factors to mediate recombination. The FLP recombinase (FLP) recognizes FRT sites and mediates recombination between them [33]. The Cre (causes recombination) recombinase recombines loxP (locus of crossover (×) in P1) target sites [34] and shares the common integration mechanism with FLP. A considerable limitation of both the FLP/FRT and the Cre/loxP system exists when used for integration and inversion. In both cases two identical target sites (homotypic sites) are present in close proximity after the recombination event, which then can serve again as substrates for a further recombination event, i.e., for excision or reinversion. In other words, FLP and Cre are "bidirectional" recombinases. Because excision reactions are kinetically favored over integrations, the integrated DNA is highly unstable [35].

The integrase of the *Streptomyces* phage PhiC31 normally mediates the integration of the phage genome [36] into the bacterial chromosome through heterotypic recombination sites, termed attB (attachment site Bacterium) and attP (attachment site Phage) sites, and like FLP and Cre, it does not need any accessory factors to mediate this integration [37]. The two sequences, though largely different, share a 3bp central region, where the crossover occurs, and this central region is flanked by imperfect repeats. Recombination between the two attachment sites creates the hybrid sites attL and attR (Left and Right) that are no longer substrates for the PhiC31 integrase, thus rendering the recombination irreversible (see Figure 4) [38][39]. This distinguishes it from Cre and FLP, and thus the PhiC31 integrase, as a unidirectional recombinase, is an efficient tool to insert transgenes into a genome [40][41]. This recombination system was successfully applied in *bacteria* [37], yeast [42], mammalian cell lines [41] [43], Xenopus [44], Drosophila [45][46][47] and plants [48][49] and so, the PhiC31-based transgenesis has become widely used. In this technique, the attB site-containing vectors are co-injected with PhiC31 integrase encoding mRNA into embryos that harbor a previously introduced transgenic attP site in their genome. The enzyme will then recombine the vector attB with the genomic attP sites, resulting in the integration of the full transgene vector into the genomic

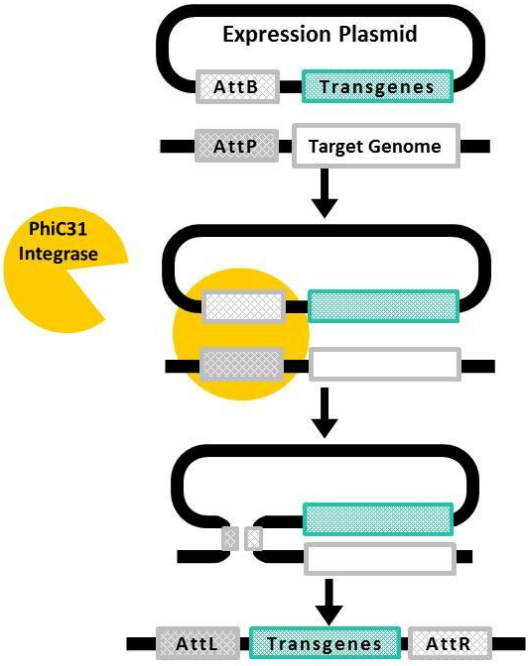



**Figure 4.** PhiC31 Integrase System. The phiC31 integrase mediates recombination between the two attachment sites. In the presence of PhiC31 integrase, an attB-containing donor plasmid can be unidirectionally integrated into a target genome through recombination at sites with sequence similarity to the native attP site.

attP locus with high efficiency, regardless of its size. The well-characterized integration sites eliminate time and effort to map transgene insertions and eliminate the need for several independent lines per transgene, thus effectively reducing the space requirement and workload of generating, analyzing, and maintaining transgenics.

Although site-specific recombination systems are diverse, the common property that makes these enzymes so attractive (the ability to specifically and autonomously integrate, excise, or invert defined sequences of DNA) also limits their practical utility. Site-specific recombinases are involved in essential biological functions, what demonstrates their strict specificity toward their natural target. Indeed, the use of these enzymes in mammalian cells requires, either the presence of rare pre-existing recognition sites or the previous introduction of specific target sites within the host genome by homologous recombination [40][50]. Therefore, besides site-specific recombinases potential as transformative tools for targeted genetic engineering, their application is conditioned by technical constraints. In order for this technology to reach its full potential, methods for the enhancement and design of custom recombinases capable of modifying specific DNA sequences are required.

#### 2.2.1.2. CRISPR/CAS9

Recently, a new class of genome engineering tools was reported and it has provided a much simpler and more economic method for gene-targeted modification, the CRISPR/Cas9 system. This follows the previous efforts in genetic manipulation, including homologous recombination [51] and RNA interference (RNAi) [52]. Other recent approaches to targeted genome modification, like zinc-finger nucleases (ZFNs) [53] and transcription-activator like effector nucleases (TALENs) [54], enable researchers to generate permanent mutations by introducing double-stranded breaks to activate repair pathways but are costly and time-consuming approaches, limiting their general use. The clustered regularly interspaced short palindromic repeats (CRISPR) are unique sequences in between DNA repeats that naturally occur and were shown to play an essential role in immunity of selected *bacteria* and *archaea*, enabling the organisms to respond to and eliminate invading genetic material. These repeats were initially discovered in the 1980s in *E. coli* [55], but their potential was only confirmed with the demonstration that *S. thermophilus* can acquire resistance against a bacteriophage by integrating a genome fragment of an infectious virus into its CRISPR locus [56].

There are three different types of CRISPR mechanisms, being type II the most studied. In this case, DNA from viruses or plasmids is cut into small fragments and integrated into a CRISPR locus among a series of short repeats. The loci are transcribed, and transcripts are then processed generating small RNAs (crRNA – CRISPR RNA), which are used to guide effector endonucleases that target invading DNA based on sequence complementarity [57]. This CRISPR mechanism distinguishes from the others as only one Cas protein (Cas9) is required for gene silencing. Cas9 (also known as Csn1), comes from Streptococcus pyogenes and has been shown, through knockdown and rescue experiments, to be a key player in this CRISPR system as it participates in the processing of crRNAs, and is responsible for the destruction of the target DNA. To achieve site-specific DNA

recognition and cleavage, Cas9 has to be complemented with a crRNA and a separate trans-activating crRNA (tracrRNA), partially complementary to the crRNA [57][58]. More recently, researchers developed a simplified two-component system that combines tracrRNA and crRNA into a single synthetic single guide RNA (sgRNA). This mechanism was shown to be as effective as the original CRISPR in guiding targeted gene alterations [57].

To date, there are three variants of the Cas9 nuclease that have been implemented in genome-editing protocols. The first is the wild-type Cas9, which can site-specifically generate a DNA double-strand break at the targeted genome locus that activates repair through error-prone nonhomologous end joining or homology-directed repair. In the absence of a template, the nonhomologous end joining is activated, resulting in insertions and/or deletions that disrupt the target loci. In the presence of a donor template with homology to the targeted locus, the homology-directed repair pathway operates, allowing for precise mutations to be made (see Figure 5A) [59][60]. The second variant has increased precision and is a Cas9 mutant form, known as Cas9D10A, with only nickase activity. This means that it cleaves only one DNA strand, and does not activate nonhomologous end joining. Instead, when provided with a homologous repair template, DNA repairs are conducted via the high-fidelity homology-directed repair pathway only, resulting in reduced insertion and/or deletion mutations (see Figure 5B) [57][61]. The last variant is a nuclease-deficient Cas9 (dCas9) with mutations in two specific Cas9 domains that inactivate cleavage activity, but do not prevent DNA binding and therefore, this variant can be used to sequence-specifically target any region of the genome without cleavage [62][63], as a gene silencing or activation tool by fusing with various effector domains [64][65] and as a visualization tool, for instance, the fusion with enhanced green fluorescent protein (EGFP) allowed visualizing repetitive DNA sequences with a single sgRNA or nonrepetitive loci using multiple sgRNAs (see Figure 5C) [66].

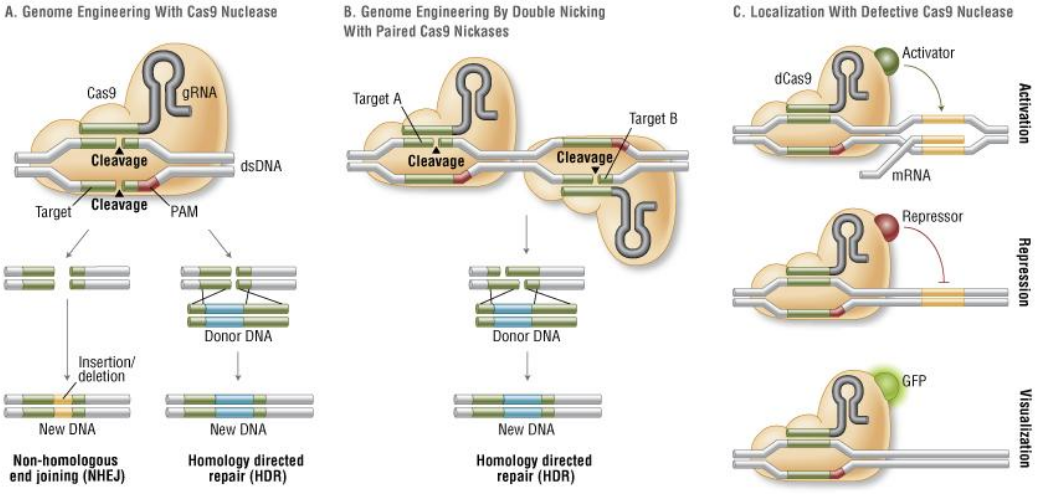



**Figure 5.** CRISPR/Cas9 genome engineering with Cas9 nuclease variants. (A) Wild-type Cas9 nuclease site specifically cleaves double-stranded DNA. Non-homologous end joining can result in insertions/deletions disrupting the target sequence. Alternatively, precise mutations and knock-ins can be made by providing a homologous repair template and exploiting the homology directed repair pathway. (B) Mutated Cas9 makes a site specific single-strand nick. Two sgRNA can be used to introduce a staggered double-stranded break which can then undergo homology directed repair. (C) Nuclease-deficient Cas9 can be fused with various effector domains allowing specific localization. For example, transcriptional activators, repressors, and fluorescent proteins. Image adapted from [83].

Besides being a faster alternative for cell genome manipulation compared to traditional gene targeting methods, the CRISPR/Cas9 system has some limitations. First, there is the possibility of off-site effects, when a mutation is introduced at non-specific loci with similar, but not identical, homology to the target sites. These can be difficult to identify and require scanning the genome for mutations at sites with sequence similarity to the sgRNA target sequence. Still, this mechanism has better targeting efficiency comparing with the previously established methods (TALENs or ZFNs) that, in human cells, were only able to achieve efficiencies ranging from 1% to 50% [67][68]. In contrast, the Cas9 system has been reported to have efficiencies up to 70% in zebrafish and plants [69], and ranging from 2–5% in induced pluripotent stem cells [70]. In addition, other studies were able to improve genome targeting up to 78% in one-cell mouse embryos, and to achieve effective germline transmission through the use of dual sgRNAs to simultaneously target an individual gene [71]. Second, mosaicism can occur when nucleases do not cut the DNA at the one cell stage of embryonic development of a mouse line and when that happens, mice with a mutant allele in only some of their cells can be produced. Multiple alleles are another drawback of the technique, due to the fact that the healing of the nuclease cleavage site by non-homologous end joining can produce cohorts of mice with different mutations from the same targeting constructs, requiring genome sequencing to verify the nature and position of the specific mutation. The production of mice with mosaics of multiple mutations is also possible and breeding may be required to segregate and isolate mice that carry single mutations.

Despite the limitations, the CRISPR/Cas9 system has been widely adopted and successfully used to target important genes in many cell lines and organisms, including human [70], *bacteria* [72], zebrafish [73], *C. elegans* [74], plants [70], *Xenopus tropicalis* [75], yeast [76], drosophila [77], monkeys [78], rabbits [79], pigs [74], rats [80] and mice [81]. This technique requires only the redesign of the crRNA to change target specificity, contrasting with other genome editing tools, where redesign of the protein-DNA interface is required. The rapid progress in developing Cas9 into a set of tools for cell and molecular biology research has been remarkable, likely due to the simplicity, high efficiency and versatility of the system. Of the designer nuclease systems currently available for precision genome engineering, the CRISPR/Cas system is by far the most user-friendly and its potential and usefulness seems to be unlimited.

## 2.3. Purpose of This Thesis

Taking into account the difficulties that the previous established methods bring to the genome editing process, in this thesis we propose a new mechanism to insert a gene, of any size, into an embryonic stem cell in a fast and efficient manner. This will allow the easy and versatile manipulation of embryonic stem cells to create embryoid bodies and mouse lines that will posteriorly allow the study of several different mechanisms in transversal fields of biology.

# 3. MATERIALS & METHODS

## 3.1. POLYMERASE CHAIN REACTION

Polymerase chain reaction (PCR) is a cell-free technique, which is used to amplify a specific DNA sequence by *in vitro* enzymatic replication assisted by a DNA polymerase. In an exponential manner a small amount of DNA is amplified into a large amount of DNA in a very short period. The region comprised between the two selected primers is the region of interest to be amplified. The DNA containing the sequence to be amplified is incubated in a test tube with the primers, each complementary to the ends of the targeted DNA, the four deoxynucleotides and a DNA polymerase.

The PCR process consists of a series of about 25-35 subsequent cycles and each cycle consists of three precisely time- and temperature-controlled steps (denaturation, annealing and extension). The first step separates the double stranded DNA into two single strands by use of a high temperature (95°C - 98°C, depending on the manufacturers' indications). This denaturing step breaks the hydrogen bonds between the two strands. Then the temperature is lowered to the primers' specific annealing temperature to allow them to base pair to their complementary sequences on the template strands. Further the reaction is heated to 72°C, the optimal temperature for the heat stable DNA polymerase to replicate the single stranded DNA segments. The DNA polymerase uses deoxynucleotides as building blocks of the new strands.

PCRs were set up with either Phusion High-Fidelity DNA Polymerase from Thermo Scientific for DNA extraction or NZYTaq 2× Green Master Mix from NZYTech to identify positive clones using optimal reaction conditions according to the manufacturer's protocol (see list of primers in the Appendix).

**Table 1.** Component and cycling instructions for the PCR reactions. The reagents pipetting information (left) and cycling conditions (right) for both Phusion High-Fidelity DNA Polymerase (top) and NZYTaq 2× Green Master Mix (bottom). X°C represents the annealing temperature of the primers used in the reaction.

| Components | 50µL Reaction | Final Concentration |
|---|---|---|
| *H2O* | up to 50µL | |
| *5X Phusion HF Buffer* | 10µL | 1X |
| *10 mM dNTPs* | 1µL | 200µM each |
| *Forward Primer* | - | 0.5µM |
| *Reverse Primer* | - | 0.5µM |
| *Template DNA* | - | 1pg-10ng |
| *Phusion DNA Polymerase* | 0.5µL | 0.02U/µL |

| Cycle Step | Temperature | Time | Cycles |
|---|---|---|---|
| *Initial Denaturation* | 98°C | 30sec | 1 |
| *Denaturation* | 98°C | 5-10sec | |
| *Annealing* | X°C | 10-30sec | 25-35 |
| *Extension* | 72°C | 15-30sec | |
| *Final Extension* | 72°C | 5-10min | 1 |
| | 4°C | hold | |

| Components | 20µL Reaction | Final Concentration |
|---|---|---|
| *H2O* | up to 20µL | |
| *Forward Primer* | - | 0.5µM |
| *Reverse Primer* | - | 0.5µM |
| *Template DNA* | - | 0.01-0.5µg |
| *NZYTaq 2× Green Master Mix* | 10µL | |

| Cycle Step | Temperature | Time | Cycles |
|---|---|---|---|
| *Initial Denaturation* | 95°C | 120sec | 1 |
| *Denaturation* | 95°C | 30-60sec | |
| *Annealing* | X°C | 30-60sec | 25-35 |
| *Extension* | 72°C | 60sec/kb | |
| *Final Extension* | 72°C | 5-10min | 1 |
| | 4°C | hold | |

## 3.2. Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to confirm each step of the cloning process and to separate PCR products to purify the desired DNA fragments. These fragments are separated by applying an electric field to move the negatively charged molecules through an agarose matrix that are separated by size in the agarose gel matrix. Shorter nucleic acid molecules move faster and migrate further than longer ones due to the fact that these drift more easily through the pores of the gel.

The matrix used was 1% w/v% agarose dissolved in 1xTBE buffer (1M Tris-base, 1M Boric acid, 0.02M EDTA). 1kb DNA Ladder or 1kb Plus DNA Ladder from New England Biolabs was run in parallel to the samples and DNA Gel Loading Dye (6X) from Thermo Scientific was added to all samples before loading them in the gel (except in the case of reactions assembled with NZYTaq 2× Green Master Mix that may be directly loaded onto agarose gels). Electrophoresis was carried out in 1X TBE buffer at 100V for 40-60 minutes. Visualization of DNA was performed by exposure to UV-light using Chemidoc XRS+ (Bio-Rad) and the products of the reactions were confirmed by size comparison with the DNA ladder.

## 3.3. DNA Extraction from Agarose Gel and Purification

Gel extraction is used to isolate the desired fragment of DNA from the agarose gel following electrophoresis. This technique involves identifying the fragments of interest, isolating the corresponding band, isolating the DNA from it and removing the accompanying salts and stain.

The DNA fragments were visualized on a Dual-Intensity Transilluminator using the low setting of UV-light and the desired band was physically removed by cutting it out from the agarose gel with a clean scalpel. The extraction of the DNA fragments from the gel slice was performed using QIAquick® Gel Extraction Kit from Qiagen following the manufacturer's protocol. DNA was eluted from the spin column using 30µL of $dH_2O$.

## 3.4. EtOH Precipitation

Ethanol precipitation is a commonly used technique to purify and/or concentrate nucleic acids (DNA or RNA) preparations in aqueous solution.

For this procedure salt and ethanol are added to the aqueous solution (1/10 volume of sodium acetate (3M, pH 5.2) and 2.5–3.0 X volume (calculated after addition of sodium acetate) of 96% ethanol), followed by a 2 hours incubation at -20°C, which forces the precipitation of nucleic acids out of solution. After precipitation the nucleic acids can then be separated from the rest of the solution by 20 minutes centrifugation at 4°C at 13000rpm. The pellet is washed in 200µL of cold 70% ethanol then after a further 15 minutes centrifugation at 4°C at 13000rpm the ethanol is removed, and the nucleic acid pellet is air dried before being resuspended.

## 3.5. Digestion of DNA with Restriction Enzymes

Restriction enzyme digestion takes advantage of naturally occurring enzymes that recognize specific sequences in DNA and cleave it at these specific sites. Restriction digestion of plasmid constructs provides a fast and cost-efficient method of gaining indirect sequence information, giving insights for the presence or absence of an insert, orientation of the insert, plasmid size, and site-specific sequence data. This technique is also useful to produce a DNA fragment that can be cloned directly into a vector. Plasmid DNA is digested with one or more restriction enzymes selected to give a distinct DNA band pattern that is easily resolved by electrophoresis.

DNA digestions were performed using restriction enzymes from New England Biolabs® or Thermo Scientific. Restriction enzyme digestions were performed under optimal conditions described by the manufacturer (see list of all the restriction enzymes in the Appendix).

## 3.6. Ligation of DNA Fragments

The ligation of DNA fragments is accomplished by covalently connecting the insert DNA (gene or fragment of interest) into a compatibly digested vector backbone and is mediated by a DNA ligase enzyme that catalysis the formation of covalent phosphodiester linkages, which permanently join the nucleotides together. This reaction involves DNA fragments that have been generated by restriction enzyme digestion. Most restriction enzymes digest DNA asymmetrically across their recognition sequence, resulting in a single stranded overhang on the digested end of the DNA fragment. The overhangs, called sticky-ends, when compatible, meaning that the overhanging base pairs on the vector and insert are complementary, allow the vector and insert to bind to each by the ligation reaction. PCR usually generates blunt-ended products and also the digestion with some restriction enzymes, meaning that there are no overhanging base pairs in the fragment. The ligation of blunt-ended products does not involve base-pairing of the protruding ends, instead, the reaction depends on random collisions between the blunt-ends.

Ligation of DNA fragments was performed using T4 DNA Ligase from New England Biolabs® with supplemented buffer. Ligation took place under optimized reaction conditions depending on the ligation type (see Table 2).

**Table 2.** Reaction mixtures and conditions that were used for either sticky- or blunt-end ligations in the present study.

| Sticky-end Ligation | |
|---|---|
| **Components** | **20µL Reaction** |
| *Vector DNA (4kb)* | 50ng (0.020pmol) |
| *Insert DNA (1kb)* | 37.5ng (0.060pmol) |
| *10X T4 DNALigase Buffer* | 2µL |
| *Nuclease-free water* | up to 20µL |
| *T4 DNA Ligase* | 1µL |
| Incubate 1H at RT | |

| Blunt-end Ligation | |
|---|---|
| **Components** | **20µL Reaction** |
| *Vector DNA (4kb)* | 50ng (0.020pmol) |
| *Insert DNA (1kb)* | 37.5ng (0.060pmol) |
| *10X T4 DNALigase Buffer* | 2µL |
| *50% PEG 4000 solution* | 2µL |
| *Nuclease-free water* | up to 20µL |
| *T4 DNA Ligase* | 1µL |
| Incubate ON at RT | |

## 3.7. GATEWAY® TECHNOLOGY

The Gateway® Technology is a cloning system based on site-specific recombinases, providing a fast and efficient way to insert one or multiple DNA sequences into vectors. Two recombination reactions constitute the basis of this technology: the BP reaction, that allows recombination of an attB substrate (attB-PCR product or a linearized attB expression clone) with an attP substrate (donor vector) creating an attL-containing entry clone; and the LR reaction, that facilitates recombination of an attL substrate (entry clone) with an attR substrate (destination vector) thus creating an attB-containing expression clone.

To perform these recombination reactions we used Gateway® BP Clonase® II Enzyme mix and Gateway® LR Clonase® II Enzyme mix for the respective reaction and optimized the manufacturer's protocol for higher efficiency of both reactions (see Table 3).

**Table 3.** Optimized reaction mixtures and conditions for both BP and LR reactions.

| BP Reaction | | |
|---|---|---|
| **Components** | **8µL Reaction** | **Final Concentration** |
| *attB PCR Product or Linearized attB Clone* | - | 50fmol |
| *Donor Vector* | - | 150fmol |
| *TE buffer, pH 8.0* | up to 8µL | |
| *BP Clonase™ II* | 2µL | |
| Incubate ON at RT | | |
| *Proteinase K Solution* | 1µL | |
| Incubate 10 minutes at 37°C | | |

| LR Reaction | | |
|---|---|---|
| **Components** | **8µL Reaction** | **Final Concentration** |
| *5' Entry Clone* | - | 10fmol |
| *Middle Entry Clone* | - | 10fmol |
| *3' Entry Clone* | - | 10fmol |
| *Destination Vector* | - | 150fmol |
| *TE buffer, pH 8.0* | up to 8µL | |
| *LR Clonase™ II* | 2µL | |
| Incubate ON at RT | | |
| *Proteinase K Solution* | 1µL | |
| Incubate 10 minutes at 37°C | | |

## 3.8. TRANSFORMATION OF DNA INTO COMPETENT CELLS

Transformation is the process by which foreign DNA is introduced into host bacteria. The transformation process involves compromising the bacterial membrane permeability (by heat-shock, electric current, etc.), allowing the DNA to be introduced into the cytosol of the bacterium. The outgrowth of bacteria following transformation allows repair of the bacterial surface and selection of recombinant cells if the newly acquired DNA conveys antibiotic resistance to the transformed cells. Transformation of bacteria with plasmids is important for replicating plasmids.

To perform the transformation, we used 25 - 50µL (depending on the strain; 25µL One Shot® ccdB Survival™ 2 T1R Competent Cells from Invitrogen™; 50µL NZY5α Competent Cells from NZYTech) of competent cells were thawed on ice and 2-5µL of plasmid was added to the cells. Incubation of the cells continued on ice for 30 minutes and then the cells were heat pulsed at 42°C for 45 seconds followed by 2 minutes incubation on ice. 300 - 900µL of RT LB medium was added to the cells and following incubation at 37°C for 1 hour. 50/100µL of sample was plated out on agar plates with appropriate antibiotic(s) for selection. Plates were incubated at 37°C

ON. The resulting colonies were tested by PCR and the positive clones inoculated in 5 - 6mL of LB media with the correct antibiotic(s) and cultivated at 37°C in a shaker ON. Plasmids were isolated using the QIAprep® Spin Miniprep Kit from Qiage) according to manufacturer's protocol. The obtained DNA was quantified using a full-spectrum UV-Vis spectrophotometer, Nanodrop 2000, Thermo Scientific.

## 3.9. Sequencing

DNA sequencing is the process of determining the precise order of nucleotides within a DNA strand. Sequence data were obtained by LIGHTRUN sequencing, GATC biotech, of the purified plasmid DNA (5µL of DNA 80 - 100ng/µL) with the desired primer (5µL of 5µM primer).

## 3.10. Transfection of U2OS Cells

U2OS cells were routinely cultured with Dulbecco's Modified Eagle Medium (DMEM), high glucose, pyruvate, from Gilbco™, and incubated at 37°C and 5% $CO_2$ to ensure a stable environment for optimal cell growth.

Cells were transfected with Lipofectamine™ 3000 Reagent kit from Invitrogen® according to the manufacturer's protocol. In the following day, the medium was changed to eliminate dead cells and toxic reagents.

## 3.11. RNA Extraction

For RNA extraction from cells we used the RNeasy® from Qiagen and followed the manufacturer's protocol for purification of total RNA from animal cells using spin technology. We used confluent cells in a 12-well plate, disrupted in Lysis Buffer complemented with β-mercapthoethanol and homogenized. Ethanol is then added to the lysate, creating conditions that promote selective binding of RNA to the RNeasy membrane when the sample is applied to the RNeasy Mini spin column. The total RNA binds to the membrane, and contaminants are efficiently washed away, which results in high-quality RNA, eluted in 35µL of RNase-free water. All bind, wash, and elution steps are performed by centrifugation in a microcentrifuge. The RNA was then quantified using Nanodrop 2000, Thermo Scientific.

## 3.12. Real-Time PCR

To perform PCR using RNA as a starting template, the RNA must first be reverse transcribed into cDNA in a reverse transcription (RT) reaction. We first digested 1µg of the RNA samples with DNase I, RNase-free, from Thermo Scientific, following the manufacturer's protocol, to digest single- and double-stranded DNA. To inactivate DNase I we added EDTA to a final concentration of 5mM before heating for 10 minutes at 65°C.

The cDNA synthesis from RNA was performed using SuperScript® IV First-Strand Synthesis System for RT-PCR from Thermo Scientific and following the manufacturer's protocol. The cDNA samples were diluted 25 times for the subsequent quantitative real-time PCR reactions.

Quantitative real-time PCR was performed using a 7500 Fast Real-Time PCR System (Applied Biosystems®) with Power SYBR® Green PCR Master Mix from Applied Biosystems®, following the standard program. In each well of a MicroAmp® Fast Optical 96-well Reaction Plate from Applied Biosystems®, we added 5µL of cDNA sample and 15µL of qRT-PCR mixture (10µL of Power SYBR® Green PCR Master Mix, 0.5µL of forward and reverse primer's mix at 4µM and 4.5µL of Milli-Q RNase/DNase free water). The expression levels were normalized to GAPDH (a housekeeping gene).

# 4. RESULTS

We designed a PhiC31 integrase-based transgenesis system consisting of two components: the targeting vector with a genomic attP landing site and in which PhiC31 is driven by the cytomegalovirus (CMV) early enhancer/chicken beta actin (CAG) promoter and the expression vector, an attB site containing vector with an insert of interest and a fluorescent marker. When in the genome of mammalian embryonic stem cells, the targeting vector drives the expression of PhiC31. The PhiC31 attP and attB sites were designed and positioned such that PhiC31-mediated recombination displaces the PhiC31 from the CAG promoter and corresponding start codon by the in-frame insertion of the targeting sequence contained in the expression vector.

## 4.1. TARGETING VECTOR

The construction of the targeting vector involves the recombination of three entry vectors, based on the Gateway cloning system.

### 4.1.1. P5E MCS CAGGS ATTP PHIC31

The plasmid p5E MCS CAGGS AttP PhiC31 contains the CAGGS promoter, a strong synthetic promoter frequently used to drive high levels of gene expression in mammalian expression vectors, and also the AttP PhiC31 landing site that allows the integration of the expression vector to occur at this site-specific location. In order to build this plasmid, we started by digesting pCAGGS-m2G with SalI and XhoI to excise the CAGGS promoter DNA sequence. The digestion was carried out as described in the previous section (see Methods) followed by electrophoresis in 1% agarose gel and extraction of the correspondent band (1979bp) (see Figure 6A). At the same time, we digested p5E-MCS (T2K228) with XhoI and dephosphorylated the hanging ends with Calf Intestinal Alkaline Phosphatase (CIP) from New England Biolabs, following the manufacturer's protocol. CIP catalyzes the dephosphorylation of 5' and 3' ends of DNA, preventing religation of the linearized plasmid DNA. The product was then run in 1% agarose gel and the DNA extracted from the band (2810bp) (see Figure 6B). The ligation then was performed between vector and insert and the product was transformed in Kanamycin plates. The colonies that grew were digested with MluI and BamHI to confirm the insertion by electrophoresis following MiniPrep (see Figure 6C).

To introduce the attP PhiC31 landing site into the plasmid, we designed specific oligonucleotides (PhiC31 XhoI BamHI F2 and PhiC31 XhoI BamHI R2) that were mixed (5µL top and 5µL bottom oligonucleotide) and annealed using an annealing program on a PCR machine (99°C for 15 minutes and slow cooldown, -1°C/15sec until it reaches 4°C). This set of primers, when annealed, creates overhangs with XhoI and BamHI restriction sites that are relevant for an easy and correct insertion in the plasmid, between the same restriction enzymes' sites in its

sequence. For that, we digested p5E MCS CAGGs with BamHI and XhoI, run in gel and then extracted de DNA from the band (4738bp) (see Figure 6D) followed by EtOH precipitation. To set up the ligation, the vector DNA pellet was resuspended with the annealed oligonucleotides and we added 1.2µL of Ligase Buffer and 1µL of T4 DNA Ligase from New England Biolabs® and incubated at RT ON. The product was then transformed in Kanamycin plates. To confirm the insertion, the resulting clones were sequenced after MiniPrep.

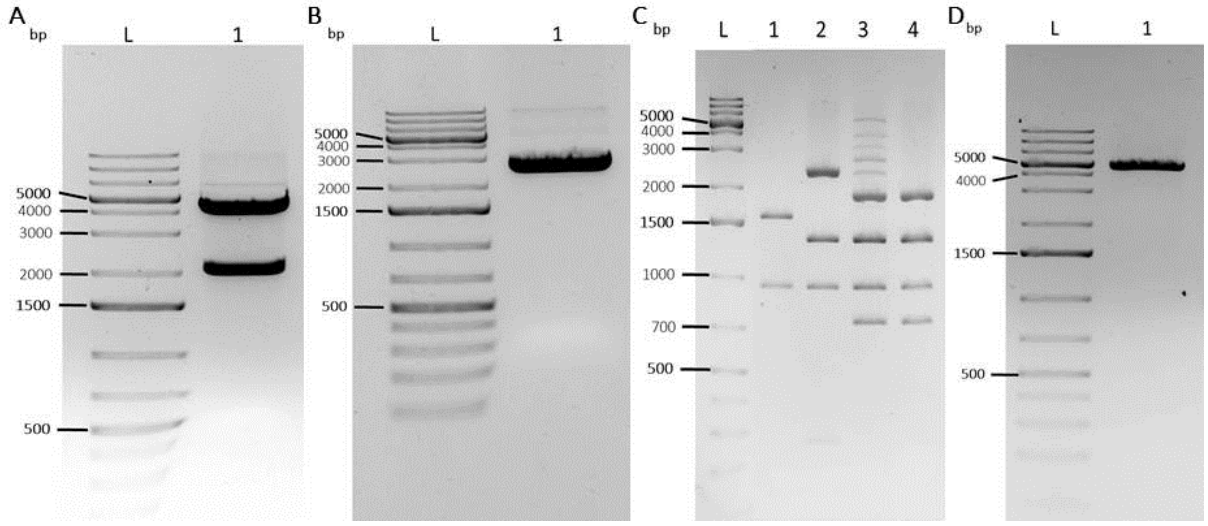



**Figure 6.** CAGGs fragment cut out from pCAGGs-m2G (RC62), linearization of p5E-MCS (T2K228), digestion to confirm CAGGs insertion in p5E-MCS (T2K228) and p5E MCS CAGGs linearization in a 1% agarose gel electrophoresis. L corresponds to the 1kb Plus DNA Ladder. (A) Band at about 2000bp in lane 1 represents the CAGGs fragment extracted by the plasmid by digestion with SalI and XhoI. (B) Band at about 2800bp in lane 1 represents linearized p5E-MCS (T2K228) after digestion with XhoI. (C) Digestion with MluI and BamHI was done to identify positive clones with expected bands of 1834, 1325, 932 and 698bp. Lanes 1 to 3 represent defective clones and lane 4 represents a positive p5E MCS CAGGs clone that was used in the following steps (D) Band at about 4700bp in lane 1 represents the linearized plasmid after digestion with BamHI and XhoI..

### 4.1.2. pME PhiC31 2A

The Middle Entry plasmid (pME PhiC31 2A) is the plasmid that contains the PhiC31 cassette and the 2A sequence. The PhiC31 cassette allows the cell to endogenously express this enzyme and the 2A sequence is a protein sequence that enables stoichiometric production of two separate proteins from the same mRNA.

This plasmid was created by amplifying a previously existing plasmid in the lab (pCS2P+PhiC31o) by PCR with the primers pME_PhiC31_F and pME_PhiC31_R to extract the PhiC31 cassette (1909bp) (see Figure 7A). The PCR product was run in gel and the DNA was extracted.

The PhiC31 fragment was then introduced in a donor vector (pDONR221 (T2K218)) by Gateway BP reaction. To confirm the insertion we digested the DNA with MluI and AgeI followed by agarose gel electrophoresis (see Figure 7B).

To introduce the 2A cassette into the plasmid, we mixed the designed oligonucleotides (2A_BglII_F and 2A_BglII_R) containing the BglII restriction site (5µL top and 5µL bottom oligonucleotide) and annealed using an annealing program on a PCR machine (99°C for 15 minutes and slow cooldown, -1°C/15sec until it reaches 4°C) . We digested pME PhiC31 with this same restriction enzyme, run in 1% agarose gel and extracted the DNA (4395bp) (see Figure 7C). We then precipitated the plasmid with EtOH and the 2A segment was introduced in the linearized plasmid by ligation. To set up the ligation, the vector DNA pellet was resuspended with the annealed oligonucleotides and we added 1.2µL of Ligase Buffer and 1µL of T4 DNA Ligase from New England Biolabs® and incubated at RT ON. After the reaction, the product was transformed and plated in Kanamycin plates and the insertion was confirmed by sequencing.

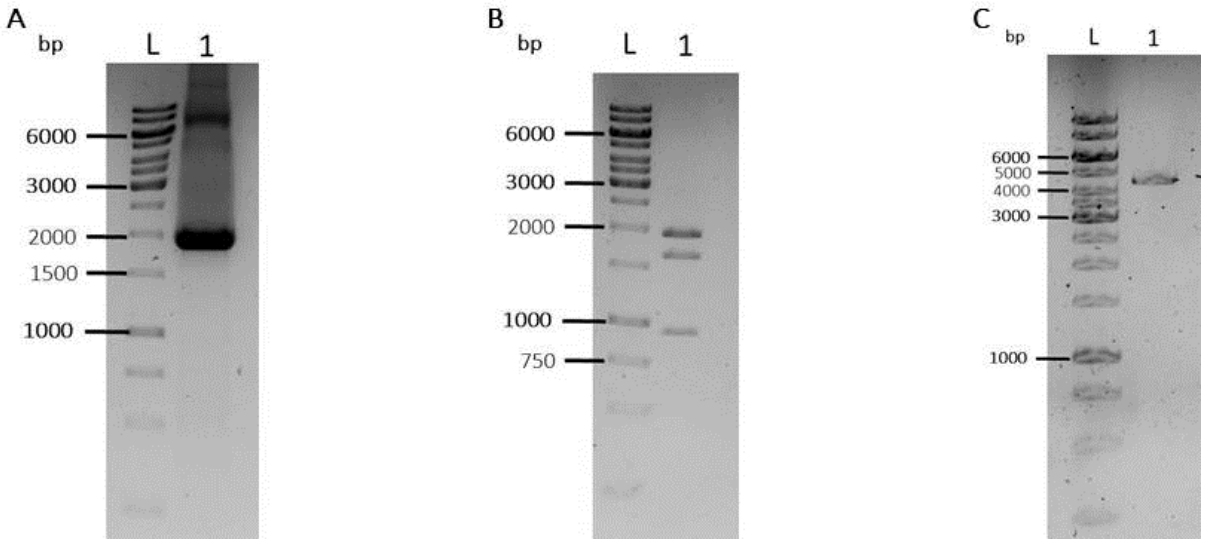



Figure 7. PhiC31 fragment cut out from pCS2P+PhiC31o, digestion of pME PhiC31 to confirm insertion and linearized pME PhiC31 in a 1% agarose gel electrophoresis. L corresponds to the 1 kb size marker. (A) Band at about 1900bp in lane 1 represents the PhiC31 fragment extracted by the pCS2P+PhiC31o plasmid by PCR with the pME_PhiC31_F and pME_PhiC31_R primers. (B) Lane 1 represents the the plasmid with the PhiC31 insertion (pME PhiC31) digested with MluI and AgeI to confirm the insertion. (C) Band at about 4400bp represents the linearized pME PhiC31 used to introduce the 2A cassette.

### 4.1.3. p3E Neo PolyA

This p3E Neo PolyA plasmid contains the Neomycin selection cassette, and the PolyA site that gives stability to the mRNA. The Neomycin antibiotic resistance sequence will allow the posterior selection of the positive clones.

This plasmid was obtained by PCR of the pROSA Dest1 to amplify the Neo PolyA with the p3E F5 Felicia and p3E R5 Felicia primers. The product was subsequently run in an agarose gel and the band was extracted according to the protocol (2164bp) (see Figure 8A).

A BP reaction was then performed between the PCR product and a donor vector (pDONRP2R-P3 (T2K220)) to create the p3E Neo PolyA plasmid that was then transformed and plated in Kanamycin plates, followed by Miniprep. The clones were confirmed by restriction enzyme digestion with BamHI and ApaLI (see Figure 8B).

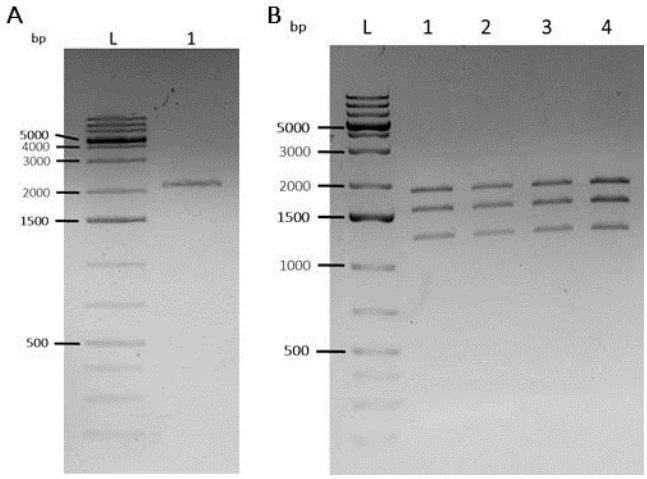



**Figure 8.** Neo PolyA fragment cut out from pROSA Dest1 and DNA digestion to confirm Neo PolyA insertion in a 1% agarose gel electrophoresis. L corresponds to the 1kb Plus DNA Ladder. (A) Band above 2000bp in lane 1 represents the Neo PolyA fragment extracted from the plasmid by PCR with the p3E F5 Felicia and p3E R5 Felicia primers. (B) Digestion with ApaLI and BamHI was done to identify positive clones with expected bands of 1880, 1598 and 1254bp. All lanes correspond to positive p3E Neo PolyA clones.

### 4.1.4. pROSA26 DV2

The transgene insertion into the ROSA26 locus by targeting has emerged as a favored solution to gain better control on gene expression. This gene is globally expressed in wild-type embryonic stem cells and does not undergo silencing, thereby representing an ideal site for directed integration of transgenes, providing high targeting efficiency and ubiquitous transgene expression at moderate levels. The drawback of this approach was the increased complexity of target vector building but to streamline the production of conditional ROSA26 targeting vectors, Nyabi et al. have generated a Gateway ROSA26 destination vector (pROSA26-DV2) as a substrate for the efficient insertion of cDNAs available as attR4-cDNA-attR2 entry clones. This vector has the ROSA26 genomic flanking arms that allow homologous recombination with the traditional cloning.

### 4.1.5. LR Reaction

Next, we performed the LR reaction between the three entry clones previously constructed and the destination vector. This reaction was successfully performed using 10fmol of each entry vector and 20fmol of the destination vector. Because the last vector is too large (>11000bp), we previously linearized it by digestion with XbaI, for more efficient recombination. The transformation was then done but, instead of LB medium, we added 250µL of SOC medium to the cells and incubated at 37°C for 3 hours. This medium is richer than LB in terms of nutrients and results in higher transformation efficiency of plasmids. Next, we did a colony PCR with the primers

pME Color-MYH9 R and pME_PhiC31_Seq3 and run in a 1% agarose gel to check for positive clones (802bp) (see Figure 9). The plasmid was then confirmed by sequencing.

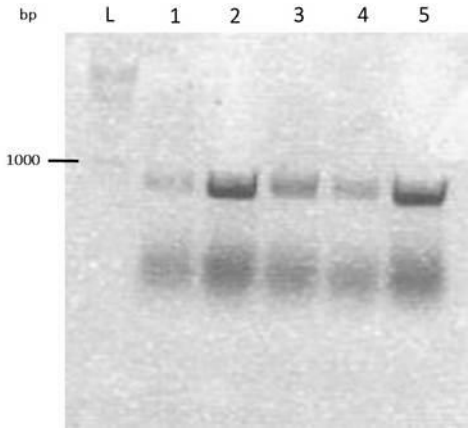



**Figure 9.** PCR to confirm LR reaction between p5E MCS CAGGs AttP PhiC31, pME PhiC31 2A, p3E Neo PolyA and pROSA26 DV2, run in a 1% agarose gel electrophoresis. L corresponds to the 1kb DNA Ladder. The band below 1000bp in each lane represent the 802bp fragment of DNA that resulted from the PCR using pME Color-MYH9 R and pME_PhiC31_Seq3 oligonucleotides to identify positive clones.

## 4.2. Expression Vector

The expression vector is the plasmid of interest that the user wants to insert in the targeting vector. To validate our strategy, we created a simple expression vector to be tested in cells. The backbone of this plasmid can be always used for the expression as it has the attachment site AttB PhiC31 that will allow it to be inserted in the targeting vector and the Puromycin selection cassette that makes possible the selection of the positive cells. The gene of interest (GOI) can be inserted in the backbone of this plasmid by a simple BP reaction.

To create this plasmid, we started by phosphorylating and annealing the AttB PhiC31 oligonucleotides (pD221-AttB site-F and pD221-AttB site-R). The phosphorylation adds a phosphate group to the 5' end of a DNA molecule making the subsequent ligation steps more efficient and, because the enzyme phosphorylates single-stranded DNA more efficiently than double stranded DNA and heat inactivation may be close to the melting temperature of the annealed oligonucleotides, this step is better performed prior to annealing the complementary oligonucleotides. The phosphorylation was performed using T4 Polynucleotide Kinase from Thermo Scientific™ and following the protocol for Phosphorylation™ of DNA. To anneal the oligonucleotides, we mixed 5µL of each phosphorylated oligonucleotide (forward and reverse) with 90µL of $dH_2O$ and used an annealing program on a PCR machine (99°C for 15 minutes and slow cooldown, -1°C/15sec until it reaches 4°C). This DNA sequence was then introduced in a donor vector (pDONR221 (T2K2018)) by cutting the plasmid with ApaI and AflII (4749bp) (see Figure 10A) and consequent ligation of the two fragments. The product was transformed in One Shot® ccdB Survival™ 2 T1R Competent Cells from Invitrogen™ in Kanamycin plates and the insertion was confirmed by PCR with M13F and ccdB Rev oligonucleotides (see Figure 10B).

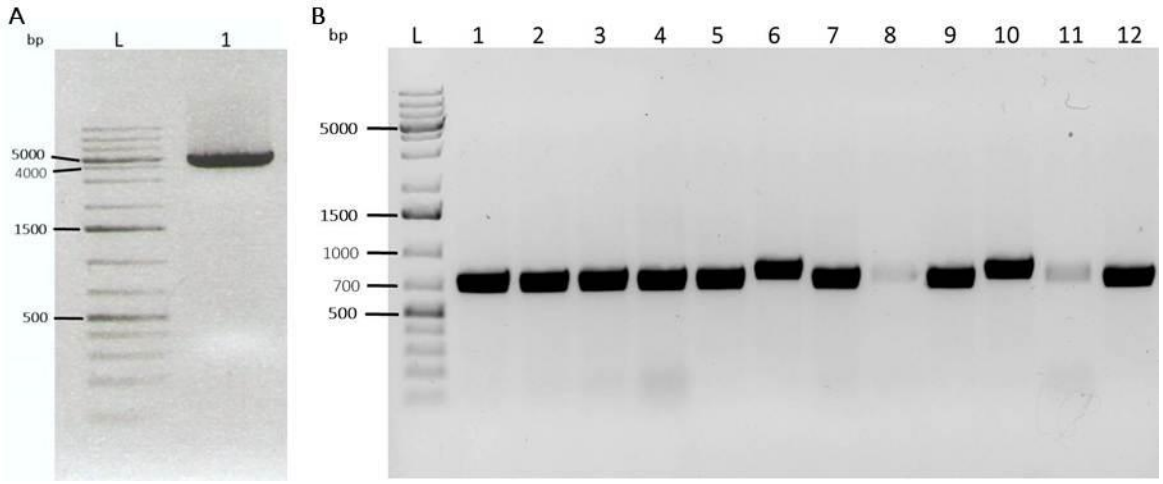



**Figure 10**. pDONR221 (T2K2018) digestion and PCR to confirm AttB PhiC31 insertion in the plasmid in a 1% agarose gel electrophoresis. L corresponds to the 1kb Plus DNA Ladder. (A) Band below 5000bp in lane 1 represents the linearized plasmid after digestion with ApaI and AflII. (B) PCR using M13F and ccdB Rev oligonucleotides to identify positive clones. Size with insert is 786bp and size without insert is 716bp. Lanes 6 and 10 correspond to positive pDONR221 AttB PhiC31 clones.

After confirming this insertion, we introduced the PGK Puro PA cassette. This sequence was amplified from a previously existing plasmid in the lab (p3E PGK Puro PA BP) by PCR and using the suitable primers (p3E-PGK-Puro-F and p3E-PGK-Puro–R). At the same time, the pDONR221 AttB PhiC31 plasmid was digested with EcoRV, that creates blunt ends, and then dephosphorylated using FastAP™ Thermosensitive Alkaline Phosphatase from Thermo Scientific™, following the manufacturer's protocol. This enzyme catalyzes the release of 5'- and 3'-phosphate groups from DNA preventing the recircularization of the cloning vector DNA during ligation. Both PCR product and linearized plasmid were run in 1% agarose gel electrophoresis and the DNA extracted (1784 and 4832bp, respectively) (see Figure 11A). The PGK Puro PA cassette was phosphorylated using T4 Polynucleotide Kinase from Thermo Scientific™ and following the manufacturer's protocol. The two fragments were then ligated, transformed in One Shot® ccdB Survival™ 2 T1R Competent Cells from Invitrogen™ and plated in Kanamycin + Chloramphenicol agar plates. The insertion was confirmed by PCR with ClamR Fw and M13R2 (see Figure 11B) and posterior sequencing, originating the backbone for the Expression Vector, named pDONR221 AttB(PhiC31) PGK PuroPA.

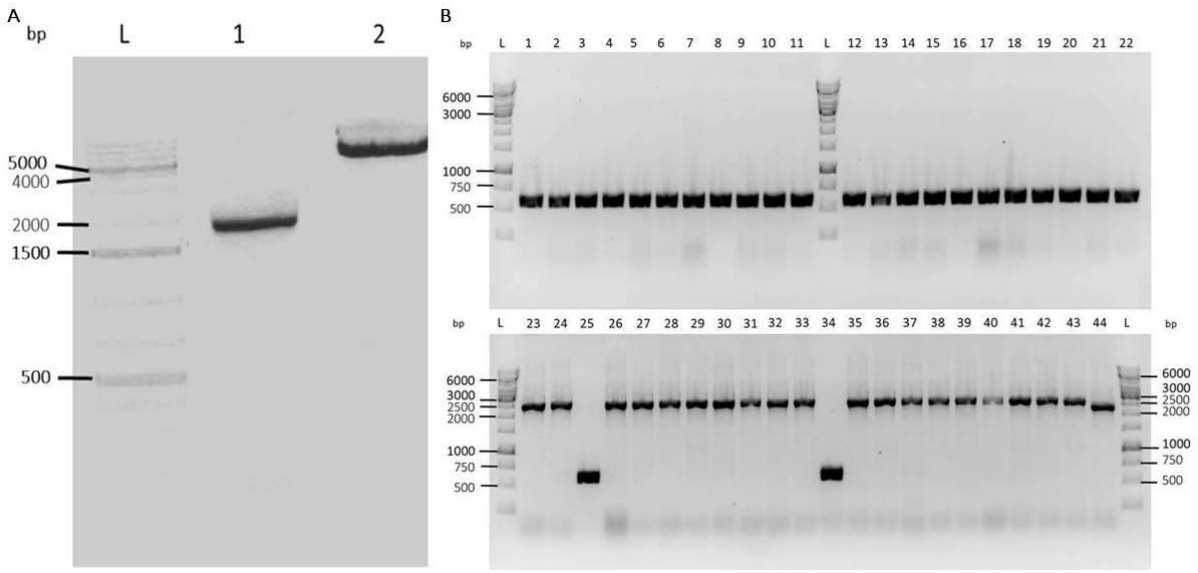



**Figure 11.** PGK Puro PA fragment cut out from p3E PGK Puro PA BP and digestion of pDONR221 AttB PhiC31 and PCR to confirm PGK Puro PA insertion in pDONR221 AttB PhiC31 plasmid in a 1% agarose gel electrophoresis. L, in A, corresponds to the 1kb Plus DNA Ladder and in B to the 1kb DNA Ladder. (A) Band at about 2000bp in lane 1 represents the PGK Puro PA fragment amplified from p3E PGK Puro PA BP plasmid by PCR with p3E-PGK-Puro-F and p3E-PGK-Puro–R oligonucleotides. Band at about 5000bp in lane 2 represents linearized pDONR221 AttB PhiC31 after digestion with EcoRV. (B) PCR was performed with ClamR Fw and M13R2 oligonucleotides. Lanes 1 to 22, 25 and 34 correspond to negative clones, which did not incorporate the PGK Puro PA fragment (590bp). Lane 44 the remaining lanes between 2000 and 2500bp, could both possible represent a positive pDONR221 AttB PhiC31 PGK Puro PA clone (2374bp). We sequenced both lane 43 and lane 44 and confirmed that the positive clone was on lane 44 being that the other lanes correspond to nonspecific bands.

As mentioned before, the following step is "user dependent" as it will allow the introduction of the GOI in the previous plasmid. Here, we inserted a GTS-mCherry.2A-H2B-eGFP cassette by amplifying by PCR a construct (pUC57-Amp_VEcad_GTS.mCherry-2A-H2B.eGFPpA_FRT.Neo.FRTpA) that was previously tested in our laboratory and that has shown positive results. By using this cassette in our plasmid, we would be able to validate its insertion into the targeting vector by accessing the fluorescence of the cells as they would express the red fluorophore in the nucleus (GTS-mCherry) and the one green one in the Golgi apparatus (H2B-eGFP), what would give us a rapid visible readout of the efficacy of our technique. In order to do this, we used designed oligonucleotides that conduct the creation of the AttB1 and AttB2 sites for the posterior BP reaction (pME-GTS.Color-F BP and pME-H2B.eGFP-R BP). The PCR product was run in 1% agarose gel electrophoresis and the DNA extracted from the band (2285bp) (see Figure 12A). The BP reaction was then performed between this product and the pDONR221 AttB PhiC31 PGK PuroPA plasmid, transformed in Kanamycin agar plates and the insertion confirmed by PCR (1117bp)(see Figure 12B) and sequencing.

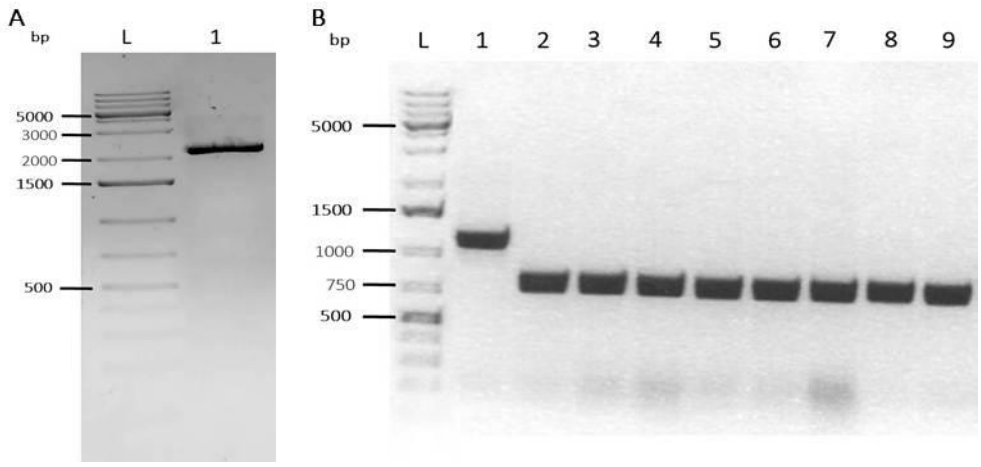



**Figure 12.** GTS Cherry H2B eGFP fragment cut out from pUC57-Amp_VEcad_GTS.mCherry-2A-H2B.eGFPpA_FRT.Neo.FRTpA and PCR to confirm GTS Cherry H2B eGFP cassette insertion in pDONR221 AttB PhiC31 PGK Puro PA plasmid in a 1% agarose gel electrophoresis. L corresponds to the 1kb Plus DNA Ladder. (A) Band above 2000bp in lane 1 represents the GTS Cherry H2B eGFP fragment amplified from pUC57-Amp_VEcad_GTS.mCherry-2A-H2B.eGFPpA_FRT.Neo.FRTpA plasmid by PCR with pME-GTS.Color-F BP and pME-H2B.eGFP-R BP oligonucleotides. (B) PCR was done with ccdB Rev, M13F and Cherry-polyA_AttP5-B2 oligonucleotides. Lanes 2 to 9 correspond to negative clones (786bp). Lane 1 represents an apparently positive pDONR221 AttB PhiC31 PGK Puro PA GTS Cherry H2B eGFP clone (1117bp).

## 4.3. Test Constructs in U2OS Cells

The human osteosarcoma U2OS cell line is one of the first generated cell lines and is widely used in various areas of biomedical research as they grow fast and have high transfection efficiency. Therefore, we used this cell line to test our constructs before using them in embryonic stem cells.

For the U2OS cells' transfection with the targeting vector, the cells were seeded in two wells in a 6 well plate 24h prior transfection and the cells from one of the wells were transfected with approximately 2µg of DNA when they reached a 30-50% confluence. The non-transfected cells served as control. We maintained the cells in culture for posterior experiments.

To check the efficiency of our strategy we did a Neomycin selection of the transfected cells in culture with DMEM complemented with G418 (500µg/µL) from Invivogen. This killed the non-transfected cells that were removed in the subsequent medium changes. When cells were confluent enough, we extracted their RNA to check for PhiC31 expression by RT-PCR using the primers qPCR PhiC31 F and qPCR PhiC31 R. We used GAPDH as housekeeping gene and the expression levels of PhiC31 in transfected and non-transfected cells were normalized to these values (see Figure 13).

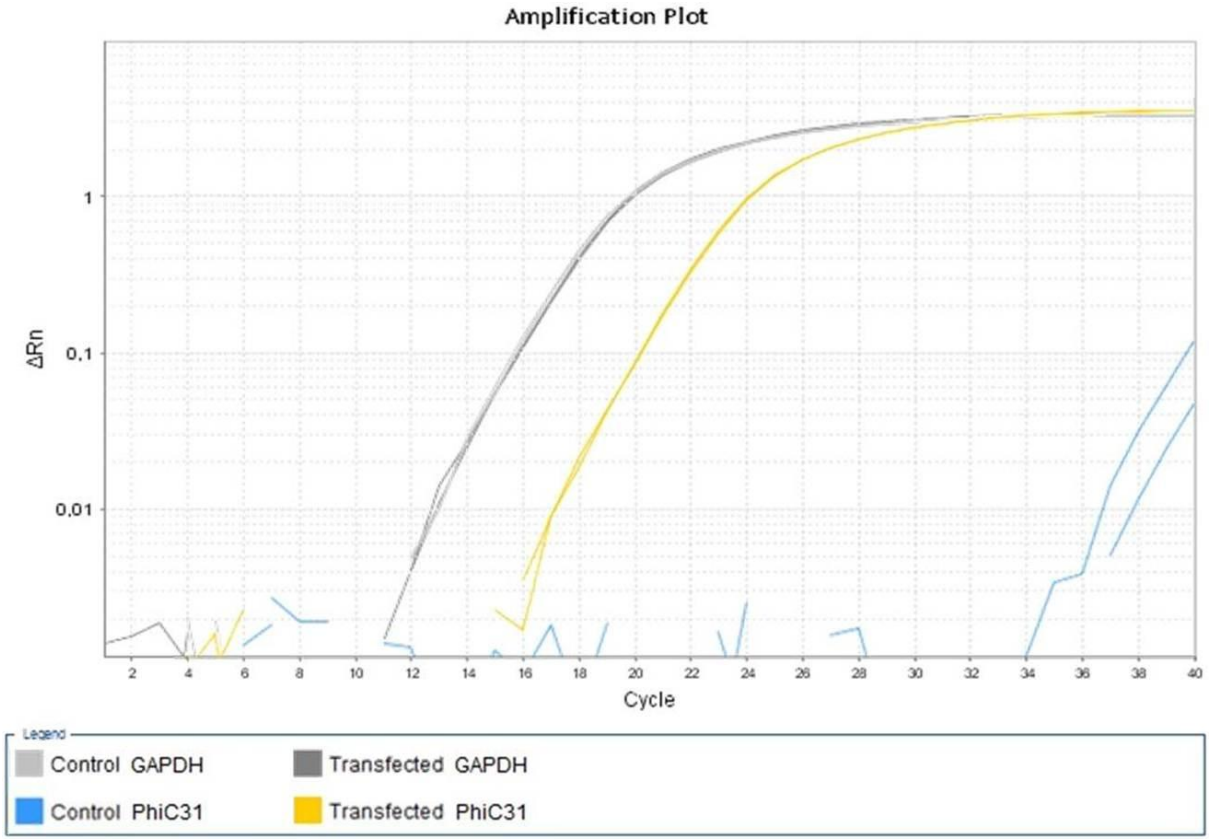



**Figure 13.** Real-time PCR plot of the cDNA of U2OS cells transfected and not transfected with the targeting vector. Gene-specific primers were designed to the hairpin of the PhiC31 mRNA precursors (qPCR PhiC31 F and qPCR PhiC31 R). The cDNA from the cells was amplified with SYBR green detection. For each condition, we made technical duplicates and the Ct values described represent the mean values. The grey plots corresponds to control (light grey, CT = 16.7) and transfected cells (dark grey, CT = 16.9) using GAPDH primers, the blue plot to the control cells with the PhiC31 primers (CT undetermined), and the yellow plot to the transfected cells also with the PhiC31 primers (CT = 21.2). CT threshold = 0.2.

The RT-PCR revealed that the control cells do not express PhiC31 (CT undetermined) contrarily to the transfected cells (CT = 21.2), confirming the efficiency of the insertion of our plasmid into the cells (see Figure 14).

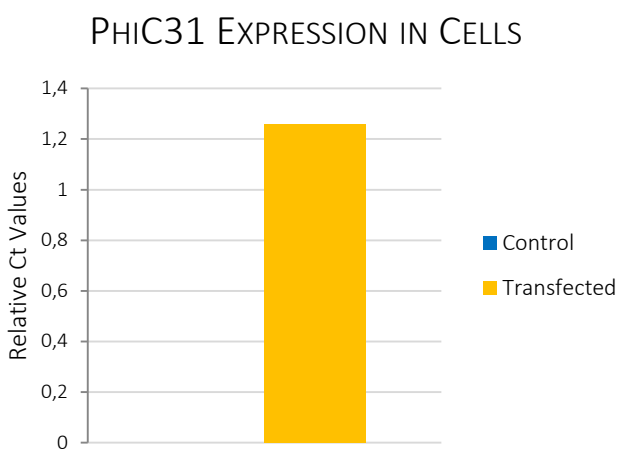



**Figure 14.** PhiC31 gene expression in Ct values relative to housekeeping gene obtained by real-time PCR. Control cells are represented in blue (not visible) and cells transfected with the targeting vector in yellow.

After confirming that our cells were expressing the enzyme, we tested the efficacy of the expression vector in these cells. In order to do this, we transfected the cells with this vector and checked for mCherry fluorescence. Both control cells and cells transfected with the targeting vector were split to 3 wells on a 24-well plate and cultured on coverslips pre-treated with 0.2% gelatin coating. For both cell "types" we created two conditions: non-transfected (control) and transfected with 500ng of DNA. Three days after transfection, cells were fixed in 1% PFA complemented with calcium chloride and magnesium chloride (0.05%), stained with DAPI, that labels the cell nuclei in blue, and GOLPH4, that labels the Golgi apparatus in green, and mounted on microscopy slides using MOWIOL with DABCO. The cells were then analyzed using a motorized widefield fluorescence microscope, Zeiss Axiovert 200M, and the MetaMorph® software (see Figure 15). We were able to see that, when transfected with the expression vector, not only the cells with the targeting vector express the mCherry fluorophore, but also the control cells (in red). We can also see that this red label is mainly located in the nuclei of the cells and not on the Golgi apparatus as we would expect.

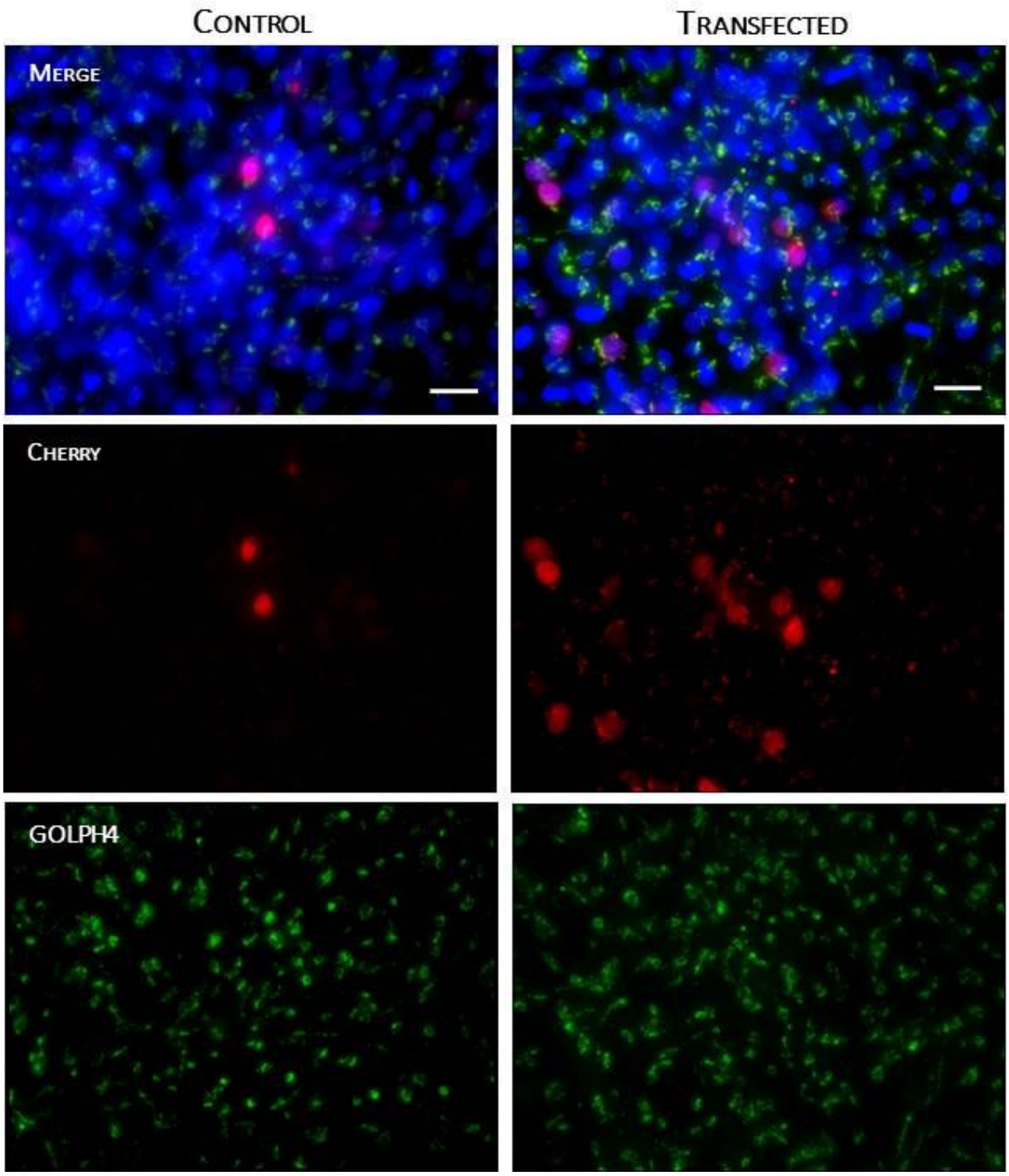



**Figure 15.** Microscopy images of the control cells (left) and cells transfected with the targeting vector (right), both transfected with the expression vector. Top images show merged channels of DAPI labelling the cell nuclei in blue, GOLPH4, staining the Golgi apparatus in green and mCherry in red, due to the transfection with the expression vector. The red fluorescence is mainly located in the nuclei of the cells but we can also see some Golgi apparatus stained with mCherry. Scale bar: 20µm.

# 5. DISCUSSION

When constructing the plasmids mentioned above, we faced some problems that slowed down the process, besides the drawbacks that are already attached to this sort of cloning methods.

The cloning of PCR amplified fragments into a plasmid vector is a routine procedure in recombinant DNA cloning, but all PCR fragments will not clone with the same efficiency and the complexity of the cloning strategy needs to be considered. The p5E CAGGS PhiC31 plasmid was the one with which we struggled the most. Our first strategy was based on the amplification of the CAGGS promoter from the pCAGGS-G2m plasmid by PCR using specific primers. This attempt failed due to the fact that the 3' part of this promoter has a high GC content and is thus refractory to PCR amplification. We then tried four different primers combinations, again, unsuccessfully. Our next step was to try the same strategy but, this time, using two other plasmids containing the promoter (pCAGGS vector in H79-1H3 and pLB2-CAG.P2G plasmid) and three other primer combinations but they were still not able to amplify the CAGGS promoter.

We then tried another PCR approach, the 'slowdown' PCR, which, theoretically, allows the amplification of extremely GC-rich (> 83%) targets, relying on combination of a novel standardized cycling protocol with varying temperature ramping rates [82]. 'Slowdown' PCR was used with a generally lowered ramp rate at 2.5°C/sec and especially a small cooling rate for reaching annealing temperature at 1.5°C/sec. It comprises an initial cycling step of 95°C for 5min for complete first-strand separation. Amplifications are run for 48 cycles with 30sec denaturation at 95°C, 30sec annealing with a progressively lowered temperature from 70°C to 53°C at a rate of 1°C every third cycle, and a primer extension of 40sec followed by 15 additional cycles with an annealing temperature of 58°C (see Figure 16).

However, despite our efforts, the promoter's amplification still did not work. Our last option was to change strategy and so we followed the "classic" cloning approach described in the previous section that worked after many attempts and optimizations.

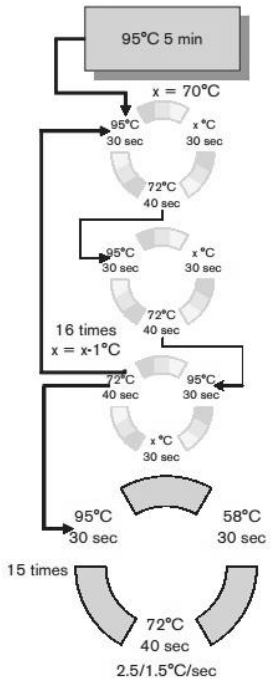



**Figure 16.** 'Slowdown' PCR cycling method. Amplifications are run for 48 cycles with 30sec denaturation at 95°C, 30sec annealing with a progressively lowered temperature from 70 to 53°C at a rate of 1°C every third cycle, and a primer extension of 40sec, followed by 15 cycles with an annealing temperature of 58°C. Ramp rates: 2.5°C/1.5°C/sec. Image adapted from [82].

Other construct that was not straightforward was the p3E Neo PolyA. Following the described strategy, our first effort to build this plasmid, in the end, revealed a problem. When designing the primers to amplify the Neo PolyA cassette we did not take into account that those oligonucleotides would lead to a segment that was not in frame. We then redesigned another set of primers and repeated the procedure.

Another step that took us a lot of time was the pROSA26-DV2 itself. This plasmid was ordered in the past and, therefore, existed in the lab for some time. When our many attempts to perform the LR reaction between it and the other three plasmids failed, we started doubting about this plasmid and we sequenced it. The sequencing results revealed a lot of mutations at important sites of this vector (namely at attachment sites) that decreased the efficiency of the reaction and that probably explained why we were not able to perform the reaction. We ordered a new plasmid but when amplifying this plasmid we faced other problem: the need of a special ccdB-resistant strain for propagation, being that the Invitrogen One Shot ccdB Survival 2 T1R Competent Cells (that we usually use in our lab) are not recommended for this plasmid and, therefore, we had to order *Escherichia coli* K12xB DB3.1 from BCCM/LMBP Plasmid Collection (Accession number LMBP 4098). We were then able to amplify the plasmid, under special conditions (30°C in Ampicillin and Chloramphenicol) and then repeat the protocol for the LR reaction. This time, after some tries and optimizations, it worked as expected.

For the expression plasmid, we projected the insertion of a GTS-mCherry.2A-H2B-eGFP cassette, which would give red fluorescence to the Golgi apparatus and green labelling to the nuclei of the cells containing the plasmid. In the last step, the BP reaction between this fragment and the pDONR221 AttB PhiC31 PGK PuroPA, it was very difficult to obtain colonies after transformation and even more difficult to get a positive clone for the insertion. In the end, we were surprised to know, by sequencing, that the H2B-eGFP segment was not inserted in the plasmid and there is no logical explanation for that to occur. Due to the difficulties that we previously encountered to successfully insert this cassette into the plasmid and the time constraints for this thesis submission, we decided to try it as it is as the GTS-Cherry fragment would be enough to prove the efficiency of our technique.

The cell transfection with the targeting vector did not give us a great input on the efficacy and efficiency of this plasmid. We confirmed that these cells expressed PhiC31 mRNA, however, the expression vector seemed to be very leaky in terms of expression and localization of the fluorescent protein. When we transfected the control cells (cells without PhiC31 expression) with the pDONR221 AttB PhiC31 GTS mCherry PGK PuroPA showed high number of cells expressing red fluorescence. Moreover, the fluorescence was supposed to be on the Golgi apparatus and not on the nuclei of the cells as we can see it. This pattern of expression was very similar to what we observed in cells expressing PhiC31. This was very surprising as we tested this expression cassette previously, and gave us the expected results (see Figure 17). This probably occurred due to an unknown defect on the plasmid, which was not integrally sequenced. The problem arouse very likely when performing the BP reaction

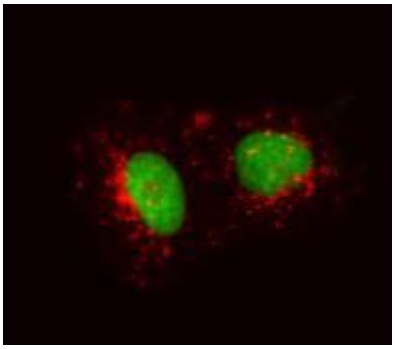


**Figure 17.** U2OS cells transfected with a construct comprising the GTS (Golgi-specific targeting sequence) fused to mCherry, a 2A self-cleavage peptide, and histone H2B fused to eGFP. We can observe the expression of mCherry in Golgi apparatus as well as eGFP inside the nucleus of these cells.

between the GTS-Cherry.2A-H2B-eGFP cassette and the correspondent donor vector. A possible explanation, which we are trying to confirm, was the insertion of the human histone H2B cassette in the same reading frame as mCherry, making the nuclei of the cells, and not the Golgi apparatus, express the red fluorophore. Given that we cannot conclude about the efficiency and efficacy of our strategy with the current expression vector, we are going to start generating a new expression vector from the beginning with a different strategy in mind.

Nevertheless, the present results with the targeting vector were satisfactory and, therefore, we believe that, in a close future, we will be able to validate our work in U2OS cells and, posteriorly apply it to embryonic stem cells. Because the technology and knowledge needed to perform this is not available in our lab, a specialized company, PolyGene Transgenetics, will ensure it.

The efficiency, ease-of-use, and versatility that can possibly be obtained with our PhiC31-based integration system represents an important advance in transgenesis approaches and opens up the possibility that will strongly facilitate such studies. Our strategy has a number of important characteristics that makes it highly promising. First, it carries out unidirectional recombination reactions, thus making the insertion of plasmid DNA efficient, permanent and stable in cells. Second, due to an endogenous source of PhiC31 integrase, we overcome the need of injecting in vitro synthesized integrase mRNA, eliminating time and costs required for mRNA production and significantly reducing complications associated with the injection process, therefore enhancing the integration rates. Third, there is no limit to the size of the plasmid to be integrated into the genome, allowing the insertion of large and/or multiple genes. Fourth, the predetermined integration sites effectively eliminate the time and effort needed to map transgene insertions. Lastly, this versatile system also offers the possibility to have an immediate visible readout to permit rapid selection for precise integration events.

Due to its high versatility, an unlimited number of constructs can be used and, therefore, we aim to build a library of expression vectors that are interesting for our lab, for example, vectors that express proteins linked to

the cytoskeleton to see the dynamic of endothelial cells in a sprouting assay and components that affect the VEGF/Notch signaling to study the competition of cells for the tip position. It would also be interesting to construct expression vectors that knockdown endogenous genes with shRNAs to generate stable knockdown cell lines avoiding the common adenoviral delivery of shRNA, which has well-known toxic effects. In our construct, after recombination, the promoter is constitutively driving the targeted protein so all the cells will be expressing it but, complementarily, it is possible to develop a vector with a temporal activation of the targeting genes by adding a Lox-Stop-Lox sequence and by crossing this mouse line with endothelial-specific Cre lines, allowing controlling both spatial and temporal activation. By adding Tamoxifen, Cre will be activated and will recombine this Lox-Stop-Lox and eliminate the Stop codon, meaning that after recombination, the cells that recombine will start to express that gene.

In summary, we strongly believe that the system presented in this study has the potential to simplify, accurate and accelerate cell genome editing processes what will allow the test of many different constructs that can be used in diverse applications. Not only by creating this embryonic stem cell line, but also, through the conception of different mouse lines, we consider that this strategy will facilitate many approaches in our laboratory and hopefully, contribute significantly to other fields of biology in a wide range of assays, allowing other researchers to optimize genome editing in their specific systems.

# REFERENCES


[1] P. Carmeliet, "Angiogenesis in Health and Disease : Therapeutic Opportunities," *Nat. Med.*, vol. 09, no. 6, pp. 653–660, 2003.

[2] J. Folkman, "Angiogenesis: an organizing principle for drug discovery?," *Nat. Rev. Drug Discov.*, vol. 6, no. 4, pp. 273–86, 2007.

[3] M. R. Swift and B. M. Weinstein, "Arterial-venous specification during development," *Circ. Res.*, vol. 104, no. 5, pp. 576–588, 2009.

[4] R. H. Adams and K. Alitalo, "Molecular regulation of angiogenesis and lymphangiogenesis," *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 6, pp. 464–478, 2007.

[5] R. Blanco and H. Gerhardt, "VEGF and Notch in tip and stalk cell selection.," *Cold Spring Harb. Perspect. Med.*, vol. 3, no. 1, p. a006569, Jan. 2013.

[6] M. Hellström, L.-K. Phng, J. J. Hofmann, E. Wallgard, L. Coultas, P. Lindblom, J. Alva, A.-K. Nilsson, L. Karlsson, N. Gaiano, K. Yoon, J. Rossant, M. L. Iruela-Arispe, M. Kalén, H. Gerhardt, and C. Betsholtz, "Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis.," *Nature*, vol. 445, no. 7129, pp. 776–80, Feb. 2007.

[7] J. D. Leslie, L. Ariza-McNaughton, A. L. Bermange, R. McAdow, S. L. Johnson, and J. Lewis, "Endothelial signalling by the Notch ligand Delta-like 4 restricts angiogenesis.," *Development*, vol. 134, no. 5, pp. 839–844, 2007.

[8] A. F. Siekmann and N. D. Lawson, "Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries.," *Nature*, vol. 445, no. 7129, pp. 781–4, Feb. 2007.

[9] K. Bentley, H. Gerhardt, and P. A. Bates, "Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation.," *J. Theor. Biol.*, vol. 250, no. 1, pp. 25–36, Jan. 2008.

[10] I. B. Lobov, R. A. Renard, N. Papadopoulos, N. W. Gale, G. Thurston, G. D. Yancopoulos, and S. J. Wiegand, "Delta-like ligand 4 (Dll4) is induced by VEGF as a negative regulator of angiogenic sprouting.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 9, pp. 3219–3224, 2007.

[11] S. Suchting, C. Freitas, F. le Noble, R. Benedito, C. Bréant, A. Duarte, and A. Eichmann, "The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 9, pp. 3225–3230, 2007.

[12] L. S. Harrington, R. C. A. Sainson, C. K. Williams, J. M. Taylor, W. Shi, J. L. Li, and A. L. Harris, "Regulation of multiple angiogenic pathways by Dll4 and Notch in human umbilical vein endothelial cells," *Microvasc. Res.*, vol. 75, no. 2, pp. 144–154, 2008.

[13] Y. Funahashi, C. J. Shawber, M. Vorontchikhina, A. Sharma, H. H. Outtz, and J. Kitajewski, "Notch regulates the angiogenic response via induction of VEGFR-1.," *J. Angiogenes. Res.*, vol. 2, p. 3, 2010.

[14] C. J. Shawber, Y. Funahashi, E. Francisco, M. Vorontchikhina, Y. Kitamura, S. A. Stowell, V. Borisenko, N. Feirt, S. Podgrabinska, K. Shiraishi, K. Chawengsaksophak, J. Rossant, D. Accili, M. Skobe, and J. Kitajewski, "Notch alters VEGF responsiveness in human and murine endothelial cells by direct regulation of VEGFR-3 expression," *J. Clin. Invest.*, vol. 117, no. 11, pp. 3369–3382, 2007.

[15] L. Jakobsson, C. A. Franco, K. Bentley, R. T. Collins, B. Ponsioen, I. M. Aspalter, I. Rosewell, M. Busse, G. Thurston, A. Medvinsky, S. Schulte-Merker, and H. Gerhardt, "Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting.," *Nat. Cell Biol.*, vol. 12, no. 10, pp. 943–953, 2010.

[16] H. Kurosawa, "Methods for inducing embryoid body formation: *in vitro* differentiation system of embryonic stem cells.," *J. Biosci. Bioeng.*, vol. 103, no. 5, pp. 389–98, 2007.

[17] A. M. Bratt-Leal, R. L. Carpenedo, and T. C. McDevitt, "Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation," *Biotechnol. Prog.*, vol. 25, no. 1, pp. 43–51, 2009.

[18] F. Y. Tsai, G. Keller, F. C. Kuo, M. Weiss, J. Chen, M. Rosenblatt, F. W. Alt, and S. H. Orkin, "An early haematopoietic defect in mice lacking the transcription factor GATA-2.," *Nature*, vol. 371, no. 6494. pp. 221–226, 1994.

[19] A. G. Elefanty, L. Robb, R. Birner, and C. G. Begley, "Hematopoietic-specific genes are not induced during *in vitro* differentiation of scl-null embryonic stem cells.," *Blood*, vol. 90, no. 4, pp. 1435–47, 1997.

[20] C. Xu, G. Liguori, E. D. Adamson, and M. G. Persico, "Specific arrest of cardiogenesis in cultured embryonic stem cells lacking Cripto-1.," *Dev. Biol.*, vol. 196, pp. 237–247, 1998.

[21] N. Ferrara, K. Carver-Moore, H. Chen, M. Dowd, L. Lu, K. S. O'Shea, L. Powell-Braxton, K. J. Hillan, and M. W. Moore, "Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene.," *Nature*, vol. 380, no. 6573. pp. 439–442, 1996.

[22] D. Vittet, M. H. Prandini, R. Berthier, a Schweitzer, H. Martin-Sisteron, G. Uzan, and E. Dejana, "Embryonic stem cells differentiate *in vitro* to endothelial cells through successive maturation steps.," *Blood*, vol. 88, no. 9, pp. 3424–3431, 1996.

[23] I. Desbaillets, U. Ziegler, P. Groscurth, and M. Gassmann, "Embryoid bodies: an *in vitro* model of mouse embryogenesis.," *Exp. Physiol.*, vol. 85, no. 6, pp. 645–651, 2000.

[24] J. Zhou, Y. Zhang, Q. Lin, Z. Liu, H. Wang, C. Duan, Y. Wang, T. Hao, K. Wu, and C. Wang, "Embryoid bodies formation and differentiation from mouse embryonic stem cells in collagen/Matrigel scaffolds," *J. Genet. Genomics*, vol. 37, no. 7, pp. 451–460, 2010.

[25] L. Jakobsson, J. Kreuger, and L. Claesson-Welsh, "Building blood vessels - stem cell models in vascular biology.," *J. Cell Biol.*, vol. 177, no. 5, pp. 751–755, 2007.

[26] L. Jakobsson, A. Domogatskaya, K. Tryggvason, D. Edgar, and L. Claesson-Welsh, "Laminin deposition is dispensable for vasculogenesis but regulates blood vessel diameter independent of flow.," *FASEB J.*, vol. 22, no. 5, pp. 1530–9, 2008.

[27] A. Fontes and U. Lakshmipathy, "Advances in genetic modification of pluripotent stem cells," *Biotechnol. Adv.*, vol. 31, no. 7, pp. 994–1001, 2013.

[28] C. Baum, O. Kustikova, U. Modlich, Z. Li, and B. Fehse, "Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors.," *Hum. Gene Ther.*, vol. 17, no. 3, pp. 253–263, 2006.

[29] M. R. Capecchi, "Altering the Genome Homologous Recombination by From ES Cells to Germ Line Chimera," *Science (80-. ).*, vol. 244, no. 4910, pp. 1288–1292, 1989.

[30] N. D. F. Grindley, K. L. Whiteson, and P. A. Rice, "Mechanisms of Site-Specific Recombination," *Annu. Rev. Biochem.*, vol. 75, no. 1, pp. 567–605, 2006.

[31] A. Landy, "Dynamic, structural, and regulatory aspects of lambda site-specific recombination," *Annu. Rev. Biochem.*, pp. 913–941, 1989.

[32] B. Sauer, "Inducible gene targeting in mice using the Cre/lox system.," *Methods A Companion to Methods Enzymol.*, vol. 14, no. 4, pp. 381–392, 1998.

[33] M. McLeod, S. Craft, and J. R. Broach, "Identification of the crossover site during FLP-mediated recombination in the Saccharomyces cerevisiae plasmid 2 microns circle.," *Mol. Cell. Biol.*, vol. 6, no. 10, pp. 3357–67, 1986.

[34] R. H. Hoess, M. Ziese, and N. Sternberg, "P1 site-specific recombination: nucleotide sequence of the recombining sites.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 79, no. 11, pp. 3398–3402, 1982.

[35] A. Baer and J. Bode, "Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for

the targeted integration of transgenes," *Curr. Opin. Biotechnol.*, vol. 12, no. 5, pp. 473–480, 2001.

[36] M. C. Smith, R. N. Burns, S. E. Wilson, and M. A. Gregory, "The complete genome sequence of the Streptomyces temperate phage straight phiC31: evolutionary relationships to other viruses.," *Nucleic Acids Res.*, vol. 27, no. 10, pp. 2145–2155, 1999.

[37] H. M. Thorpe and M. C. Smith, "*In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 10, pp. 5505–5510, 1998.

[38] H. M. Thorpe, S. E. Wilson, and M. C. M. Smith, "Control of directionality in the site-specific recombination system of the streptomyces phage PhiC31," *Mol. Microbiol.*, vol. 38, no. 2, pp. 232–241, 2000.

[39] M. C. A. Smith, R. Till, K. Brady, P. Soultanas, H. Thorpe, and M. C. M. Smith, "Synapsis and DNA cleavage in ϕC31 integrase-mediated site-specific recombination," *Nucleic Acids Res.*, vol. 32, no. 8, pp. 2607–2617, 2004.

[40] J. Bischof and K. Basler, "Recombinases and Their Use in Gene Activation, Gene Inactivation, and Transgenesis," in *Drosophila: Methods and Protocols*, 1st ed., no. 5, C. Dahmann, Ed. Humana Press, 2008, pp. 175–195.

[41] A. C. Groth, E. C. Olivares, B. Thyagarajan, and M. P. Calos, "A phage integrase directs efficient site-specific integration in human cells.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 11, pp. 5995–6000, 2000.

[42] L. C. Thomason, R. Calendar, and D. W. Ow, "Gene insertion and replacement in Schizosaccharomyces pombe mediated by the Streptomyces bacteriophage PhiC31 site-specific recombination system," *Mol. Genet. Genomics*, vol. 265, no. 6, pp. 1031–1038, 2001.

[43] B. Thyagarajan, E. C. Olivares, R. P. Hollis, D. S. Ginsburg, and M. P. Calos, "Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage ϕ C31 Integrase Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage □ C31 Integrase," vol. 21, no. 12, pp. 3926–3934, 2001.

[44] B. G. Allen and D. L. Weeks, "Transgenic Xenopus laevis embryos can be generated using ϕC31 integrase," *Nat. Methods*, vol. 2, no. 12, pp. 975–979, 2005.

[45] A. C. Groth, M. Fish, R. Nusse, and M. P. Calos, "Construction of Transgenic Drosophila by Using the Site-Specific Integrase from Phage PhiC31," *Genetics*, vol. 166, no. 4, pp. 1775–1782, 2004.

[46] K. J. T. Venken, Y. He, R. a Hoskins, and H. J. Bellen, "P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster.," *Science*, vol. 314, no. 5806, pp. 1747–1751, 2006.

[47] J. Bischof, R. K. Maeda, M. Hediger, F. Karch, and K. Basler, "An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 9, pp. 3312–7, 2007.

[48] M. Rubtsova, K. Kempe, A. Gils, A. Ismagul, J. Weyen, and M. Gils, "Expression of active Streptomyces phage phiC31 integrase in transgenic wheat plants," *Plant Cell Rep.*, vol. 27, no. 12, pp. 1821–1831, 2008.

[49] J. G. Thomson, R. Chan, R. Thilmony, Y.-Y. Yau, and D. W. Ow, "PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome.," *BMC Biotechnol.*, vol. 10, no. 1, p. 17, 2010.

[50] T. Gaj, S. J. Sirk, and C. F. Barbas, "Expanding the scope of site-specific recombinases for genetic and metabolic engineering," *Biotechnol. Bioeng.*, vol. 111, no. 1, pp. 1–15, 2014.

[51] M. R. Capecchi, "Gene targeting in mice : functional analysis of the mammalian genome for the twenty-first century," *Nat. Rev. Genet.*, vol. 6, no. June, pp. 507–512, 2005.

[52] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, "Potent and specific genetic

interference by double-stranded RNA in Caenorhabditis elegans," *Nature*, vol. 391, no. 6669, pp. 806–811, 1998.

[53] J. C. Miller, F. D. Urnov, Y. Lee, J. Rock, T. Sorba, S. Patterson, P. D. Gregory, E. J. Rebar, and M. C. Holmes, "Development of Zinc Finger Nucleases for Therapeutic Gene Correction of Sickle Cell Anemia," in *Molecular Therapy*, vol. 11, no. S1, 2005, pp. 35–35.

[54] C. Mussolino, R. Morbitzer, F. Lütge, N. Dannemann, T. Lahaye, and T. Cathomen, "A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity," *Nucleic Acids Res.*, vol. 39, no. 21, pp. 9283–9293, 2011.

[55] Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, and A. Nakata, "Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.," *J. Bacteriol.*, vol. 169, no. 12, pp. 5429–33, 1987.

[56] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, Patrick Boyaval, S. Moineau, D. Romero, and P. Horvath, "CRISPRProvides Acquired Resistance Against Viruses in Prokaryotes," *Science (80-. ).*, vol. 315, no. 2007, pp. 1709–1712, 2007.

[57] M. Jinek Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity.," *Science (80-. ).*, vol. 337, no. 6096, pp. 816–821, 2012.

[58] E. Deltcheva, K. Chylinski, C. M. Sharma, and K. Gonzales, "CRISPR RNA maturation by trans -encoded small RNA and host factor RNase III," *Nature*, vol. 471, no. 7340, pp. 602–607, 2011.

[59] S. Overballe-Petersen, K. Harms, L. a a Orlando, J. V. M. Mayar, S. Rasmussen, T. W. Dahl, M. T. Rosing, A. M. Poole, T. Sicheritz-Ponten, S. Brunak, S. Inselmann, J. de Vries, W. Wackernagel, O. G. Pybus, R. Nielsen, P. J. Johnsen, K. M. Nielsen, E. Willerslev, and J. De Vries, "Bacterial natural transformation by highly fragmented and damaged DNA.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 49, pp. 19860–5, 2013.

[60] C. Gong, P. Bongiorno, A. Martins, N. C. Stephanou, H. Zhu, S. Shuman, and M. S. Glickman, "Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C," *Nat Struct Mol Biol*, vol. 12, no. 4, pp. 304–312, 2005.

[61] L. Davis and N. Maizels, "Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 10, pp. E924–32, 2014.

[62] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim, "Repurposing CRISPR as an RNA-Guided Platform for Sequence- Specific Control of Gene Expression," *Cell*, vol. 152, no. 5, pp. 1173–1183, 2013.

[63] G. Gasiunas, R. Barrangou, P. Horvath, and V. Siksnys, "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 39, pp. E2579–86, 2012.

[64] M. L. Maeder, S. J. Linder, V. M. Cascio, Y. Fu, Q. H. Ho, and J. K. Joung, "CRISPR RNA-guided activation of endogenous human genes," *Nat Methods*, vol. 10, no. 10, pp. 977–979, 2013.

[65] P. Perez-Pinera, D. D. Kocak, C. M. Vockley, A. F. Adler, A. M. Kabadi, L. R. Polstein, P. I. Thakore, K. A. Glass, D. G. Ousterout, K. W. Leong, F. Guilak, G. E. Crawford, T. E. Reddy, and C. A. Gersbach, "RNA-guided gene activation by CRISPR-Cas9-based transcription factors," *Nat Methods*, vol. 10, no. 10, pp. 973–976, 2013.

[66] B. Chen, L. a Gilbert, B. a Cimini, J. Schnitzbauer, W. Zhang, G. Li, J. Park, E. H. Blackburn, J. S. Weissman, and S. Lei, "Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System," *Cell*, vol. 155, no. 7, pp. 1479–1491, 2013.

[67] J. C. Miller, S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley,

G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory, and E. J. Rebar, "A TALE nuclease architecture for efficient genome editing," *Nat. Biotechnol.*, vol. 29, no. 2, pp. 143–148, 2011.

[68] M. L. Maeder, S. Thibodeau-beganny, A. Osiak, D. A. Wright, R. M. Anthony, M. Eichtinger, T. Jiang, J. E. Foley, R. J. Winfrey, J. A. Townsend, E. Unger-wallace, J. D. Sander, F. Fu, J. Pearlberg, C. Göbel, J. P. Dassie, M. H. Porteus, D. C. Sgroi, A. J. Iafrate, P. B. M. Jr, T. Cathomen, D. F. Voytas, and J. K. Joung, "Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification," *Mol. Ther.*, vol. 31, no. 2, pp. 294–301, 2008.

[69] Z. Feng, B. Z. W. Ding, and X. L. D.-L. Y. P. W. F. C. S. Z. F. Z. Y. M. J.-K. Zhu, "Efficient genome editing in zebrafish using a CRISPR-Cas system," *Nat Biotechnol*, vol. 31, no. 3, pp. 227–229, 2013.

[70] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church, "RNA-Guided Human Genome Engineering via Cas9 Prashant," *Science (80-. ).*, vol. 339, no. 6121, pp. 823–826, 2013.

[71] J. Zhou, J. Wang, B. Shen, L. Chen, Y. Su, J. Yang, W. Zhang, X. Tian, and X. Huang, "Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting," *FEBS J.*, vol. 281, no. 7, pp. 1717–1725, 2014.

[72] L. Fabre, S. Le Hello, C. Roux, S. Issenhuth-Jeanjean, and F. X. Weill, "CRISPR Is an Optimal Target for the Design of Specific PCR Assays for Salmonella enterica Serotypes Typhi and Paratyphi A," *PLoS Negl. Trop. Dis.*, vol. 8, no. 1, p. 14, 2014.

[73] W. Y. Hwang, Y. Fu, D. Reyon, M. L. Maeder, P. Kaini, J. D. Sander, J. K. Joung, R. T. Peterson, and J. R. J. Yeh, "Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System," *PLoS One*, vol. 8, no. 7, pp. 1–9, 2013.

[74] T. Hai, F. Teng, R. Guo, W. Li, and Q. Zhou, "One-step generation of knockout pigs by zygote injection of CRISPR/Cas system.," *Cell Res.*, vol. 24, no. 3, pp. 372–5, 2014.

[75] X. Guo, T. Zhang, Z. Hu, Y. Zhang, Z. Shi, Q. Wang, Y. Cui, F. Wang, H. Zhao, and Y. Chen, "Efficient RNA/Cas9-mediated genome editing in Xenopus tropicalis.," *Development*, vol. 141, no. 3, pp. 707–14, 2014.

[76] J. E. DiCarlo, J. E. Norville, P. Mali, X. Rios, J. Aach, and G. M. Church, "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems," *Nucleic Acids Res.*, vol. 41, no. 7, pp. 4336–4343, 2013.

[77] S. J. Gratz, F. P. Ukken, C. D. Rubinstein, G. Thiede, L. K. Donohue, A. M. Cummings, and K. M. Oconnor-Giles, "Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila," *Genetics*, vol. 196, no. 4, pp. 961–971, 2014.

[78] Y. Niu, B. Shen, Y. Cui, Y. Chen, J. Wang, L. Wang, Y. Kang, X. Zhao, W. Si, W. Li, A. P. Xiang, J. Zhou, X. Guo, Y. Bi, C. Si, B. Hu, G. Dong, H. Wang, Z. Zhou, T. Li, T. Tan, X. Pu, F. Wang, S. Ji, Q. Zhou, X. Huang, W. Ji, and J. Sha, "Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos," *Cell*, vol. 156, no. 4, pp. 836–843, 2014.

[79] D. Yang, J. Xu, T. Zhu, J. Fan, L. Lai, J. Zhang, and Y. E. Chen, "Effective gene targeting in rabbits using RNA-guided Cas9 nucleases," *J. Mol. Cell Biol.*, vol. 6, no. 1, pp. 97–99, 2014.

[80] Y. Ma, X. Zhang, B. Shen, Y. Lu, W. Chen, J. Ma, L. Bai, X. Huang, and L. Zhang, "Generating rats with conditional alleles using CRISPR/Cas9.," *Cell Res.*, vol. 24, no. 1, pp. 122–5, 2014.

[81] A. Yasue, S. N. Mitsui, T. Watanabe, T. Sakuma, S. Oyadomari, T. Yamamoto, S. Noji, T. Mito, and E. Tanaka, "Highly efficient targeted mutagenesis in one-cell mouse embryos mediated by the TALEN and CRISPR/Cas systems," *Sci. Rep.*, vol. 4, p. 5705, 2014.

[82] H. S. Bachmann, W. Siffert, and U. H. Frey, "Successful amplification of extremely GC-rich promoter regions using a novel 'slowdown PCR' technique.," *Pharmacogenetics*, vol. 13, no. 12, pp. 759–766,

2003.

[83] A. Reis, B. Hornblower, R. Brett, and G. Tzerzinis, “CRISPR / Cas9 and Targeted Genome Editing : A New Era in Molecular Biology,” *New Engl. BioLabs Expressions*, no. l, pp. 1–4, 2014.

# APPENDIX

## PRIMERS

**Table A1.** Sequences of all primers used for either PCR amplifications or sequencing analysis. Primers used for sequencing are marked with *. All primers were imported from Sigma Aldrich®.

| Primer Name | Tm (°C) | Sequence (5' → 3') |
|---|---|---|
| 2A_BglII_F | 95,6 | GATCGGAGGGCAGGGGAAGTCTTCTAACATGCGGGGACGTGGAGGAAAATCCCGGGCCCA |
| 2A_BglII_R | 94,6 | GATCTGGGCCCGGGATTTTCCTCCACGTCCCCGCATGTTAGAAGACTTCCCCTGCCCTCC |
| CAGGs PhiC31 Seq1* | 58,7 | TATGGAGTTCCGCGTTAC |
| CAGGs PhiC31 Seq2* | 58,6 | CGTTCTGCTTCACTCTCC |
| CAGGs PhiC31 Seq3* | 62,3 | TGGTTTAATGACGGCTCG |
| CAGGs PhiC31 Seq4* | 60,7 | GAAGAAGGCATGAACATGG |
| Caggs Seq1* | 61,3 | AACCGTATTACCGCCTTTG |
| Caggs Seq2* | 66,5 | CGTAATGGCTGGCCTGTTG |
| Caggs Seq3* | 58,9 | GTGTCTTACCGGGTTGG |
| ccdB Rev* | 63,0 | CCACCGCGAAAATGACATC |
| Cherry-polyA_AttP5-B2* | 64,7 | TGATGTTGACGTTGTAGGCG |
| ClamR Forw* | 59,0 | CGATGCCATTGGGATATATC |
| GAPDH_Fw | 66,5 | GTCAAGGCTGAGAACGGGAA |
| GAPDH_Rv | 64,5 | TGGACTCCACGACGTACTCA |
| M13F* | 61,8 | TGTAAAACGACGGCCAGT |
| M13R2* | 62,1 | CAGGAAACAGCTATGACCATG |
| Neo pA Seq1* | 62,8 | TCGACGTTGTCACTGAAGC |
| Neo pA Seq2* | 61,2 | CGCCTTCTTGACGAGTTC |
| Neo pA Seq3* | 60,5 | GTCGCAGATCCAGACATG |
| p3E PGK-Puro-pA F | 83,2 | CGATCTGGATCCGCTCAGACGAGTCGGATCTCC |
| p3E PGK-Puro–pA R | 77,9 | GCTGATGGATCCAGGATAGCTAGAGCCAGACATG |
| p3E Seq Neo F* | 61,5 | ATGAACTGCAGGACGAGG |
| p3E-F5 Felicia | 88,1 | GGGGACAGCTTTCTTGTACAAAGTGGGCCAATATGGGATCGGCCATTGA |
| p3E-R5 Felicia | 82,2 | GGGGACAACTTTGTATAATAAAGTTGGGTTCCGGATCAGCTTGATGG |
| p5E CAGGs Seq2 CF* | 64,2 | AAAGCCTTAAAGGGCTCCG |
| p5E CAGGs Seq3 CF* | 63,5 | TGCCTTTTATGGTAATCGTGC |
| pD221-AttB site–F | 66,5 | TTAAGTACTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTCGGC |
| pD221-AttB site–R | 68,0 | GAGATGGGTGAGGTGGAGTACGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACCCGCACCGCGGCTTCGAGTAC |
| PhiC31 XhoI-BamHI F2 | 89,1 | TCGAGGCCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGAG |
| PhiC31 XhoI-BamHI R2 | 88,2 | GATCCTCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGGCC |
| pME Color-MYH9 R | 81,3 | AGAAAGCTGGGTTGAACCTGCGTGCAATCCATC |
| pME PhiC31 Seq1* | 67,0 | GTGTCCCGGTTCAGCAGG |
| pME PhiC31 Seq3* | 64,0 | ATCACCGGCCTGTGTAAGAG |
| pME PhiC31 Seq4* | 64,4 | CAAGAGGGTGTTCGTGGG |
| pME_PhiC31_F | 87,1 | GGGGACAAGTTTGTACAAAAAAGCAGGCTGCCGCCACCATGGATACCTA |
| pME_PhiC31_R | 83,3 | GGGGACCACTTTGTACAAGAAAGCTGGGTAGATCTCACTTTCCGCTTTTTCTTAG |
| pME-GTS.Color-F BP | 85,8 | GGGGACAAGTTTGTACAAAAAAGCAGGCTTACCAGATCAGCTCCTCCACGAAG |
| pME-H2B.eGFP-R BP | 85,0 | GGGGACCACTTTGTACAAGAAAGCTGGGTTAGGGTGAACATACGCGTCTT |
| qPCR PhiC31 F | 63,7 | CGTGATGACCAGCAAGAGAG |
| qPCR PhiC31 R | 66,5 | CGGCCATGCTCACATTACAG |

## Restriction Enzymes

**Table A2.** Recognition sequences of all restriction enzymes used in this study.

| Enzyme | Recognition Sequence | Buffer | Vendor |
|---|---|---|---|
| AflII | C/TTAAG | Orange | Thermo Scientific |
| AgeI | A/CCGGT | NEBuffer 1.1 | New England BioLabs |
| ApaI | GGGCC/C | CutSmart Buffer | New England BioLabs |
| ApaLI | G/TGCAC | Tango | Thermo Scientific |
| BamHI | G/GATCC | Buffer BamHI | Thermo Scientific |
| BglII | A/GATCT | Orange | Thermo Scientific |
| EcoRV | GAT/ATC | NEBuffer 3.1 | New England Biolabs |
| MluI | A/CGCGT | NEBuffer 3.1 | New England BioLabs |
| SalI | G/TCGAC | NZY Buffer A | NZY Tech |
| XhoI | C/TCGAG | Red | Thermo Scientific |

## CLONING PROCESS OVERVIEW

### P5E MCS CAGGS ATTP PHIC31

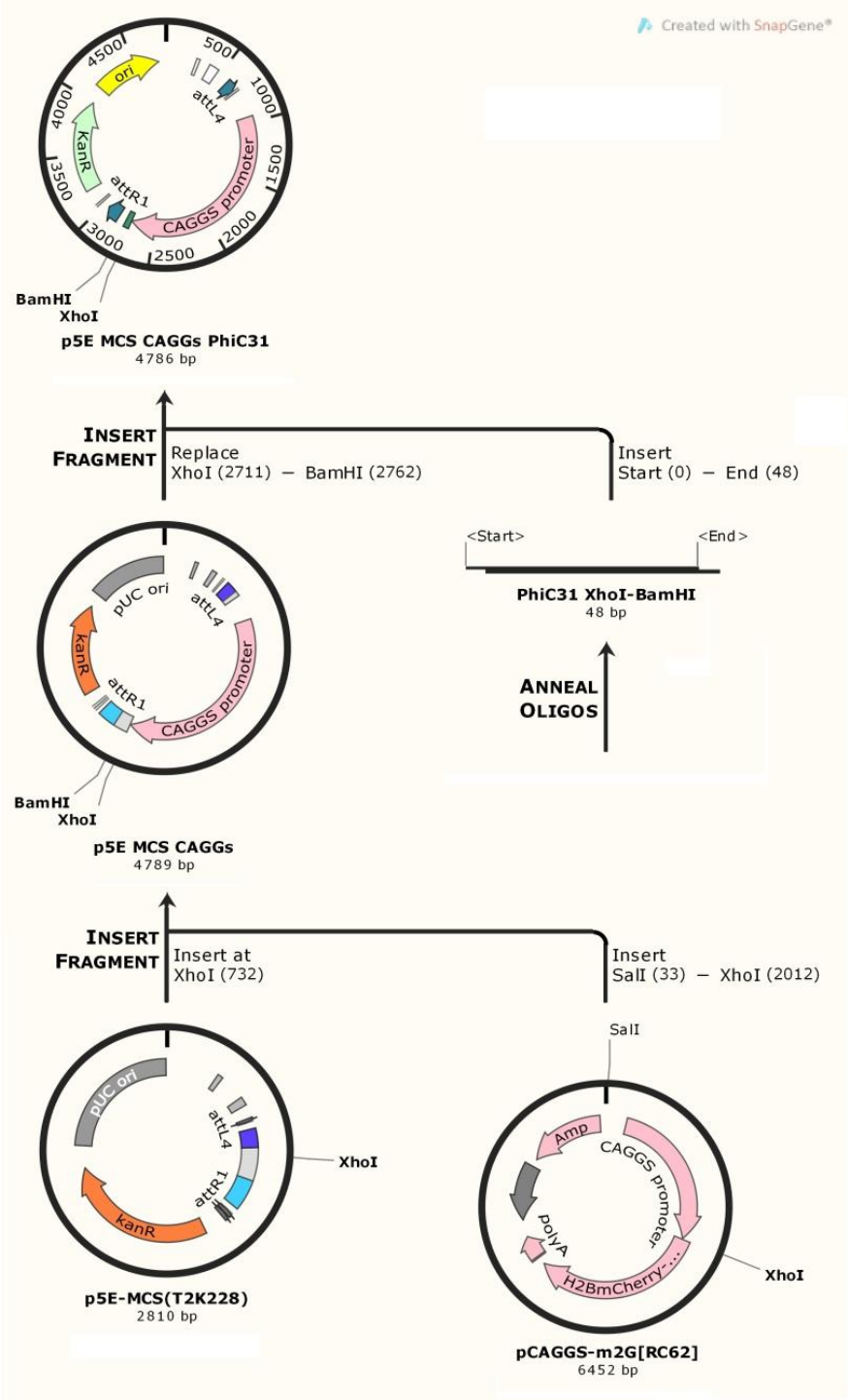



**Figure A1.** Schematic overview of the cloning process that resulted in the generation of the p5E MCS CAGGs AttP PhiC31 plasmid. The figure shows the different vectors used and the strategy followed.

pME PhiC31 2A

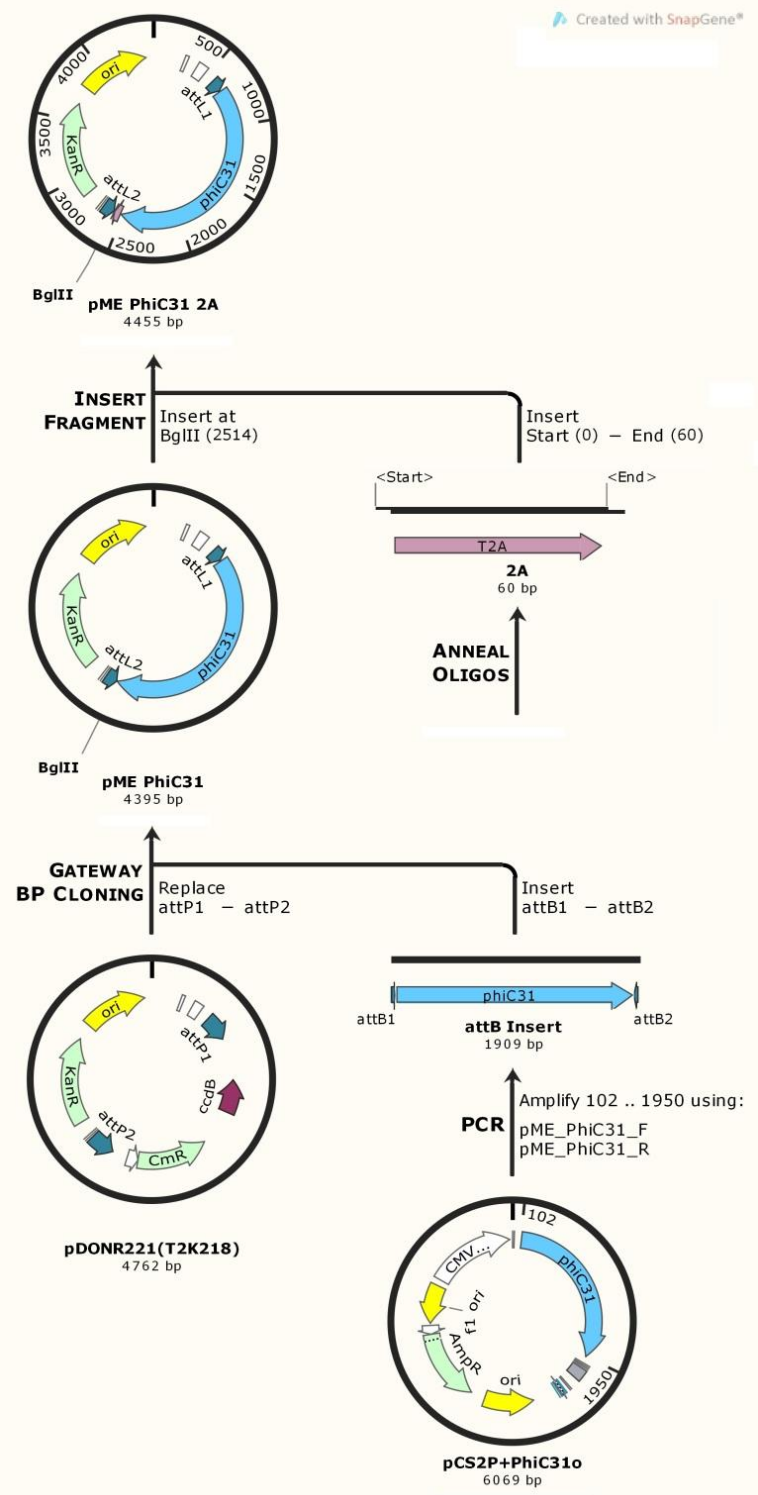



**Figure A2.** Schematic overview of the cloning process that resulted in the generation of the pME PhiC31 2A plasmid. The figure shows the different vectors used and the strategy followed.

P3E NEO POLYA

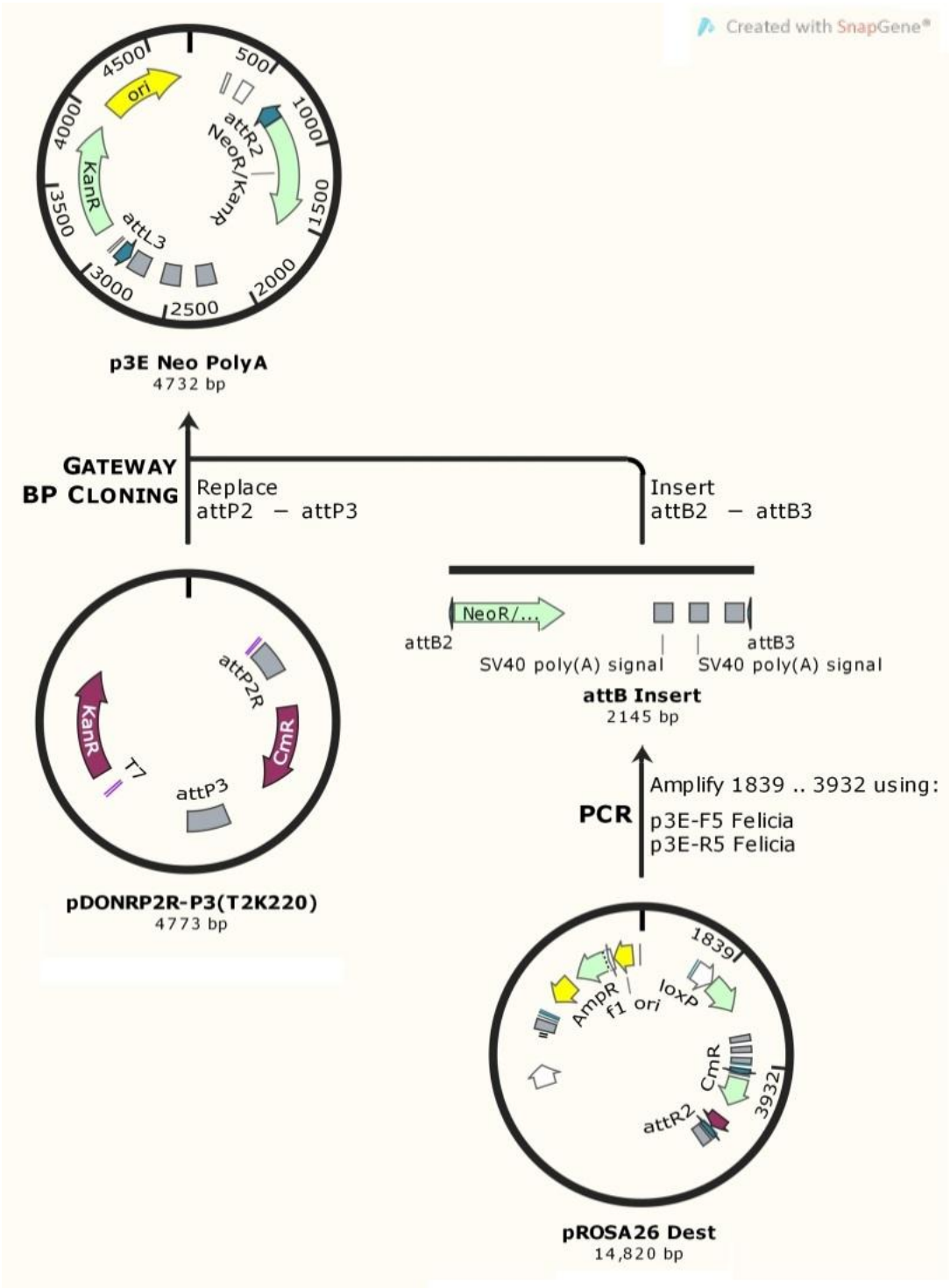



**Figure A3.** Schematic overview of the cloning process that resulted in the generation of the p3E Neo PolyA plasmid. The figure shows the different vectors used and the strategy followed.

pROSA26-DV2-CAGGS-PHIC31-NEOPA

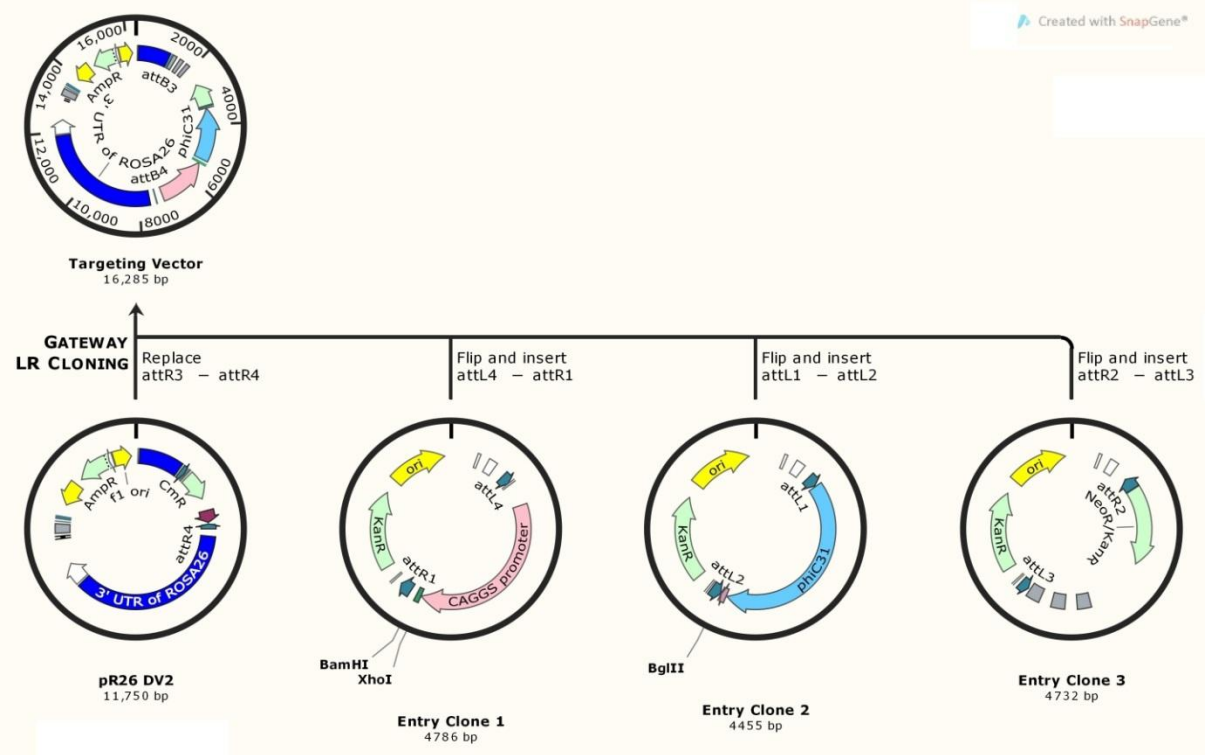



Figure A4. Schematic overview of the LR reaction (Gateway) between the three created entry vectors and the destination vector (pROSA26 DV2) that resulted in the generation of the targeting vector.

## pDONR221 AttB PhiC31 PGK Puro PA GTS Cherry

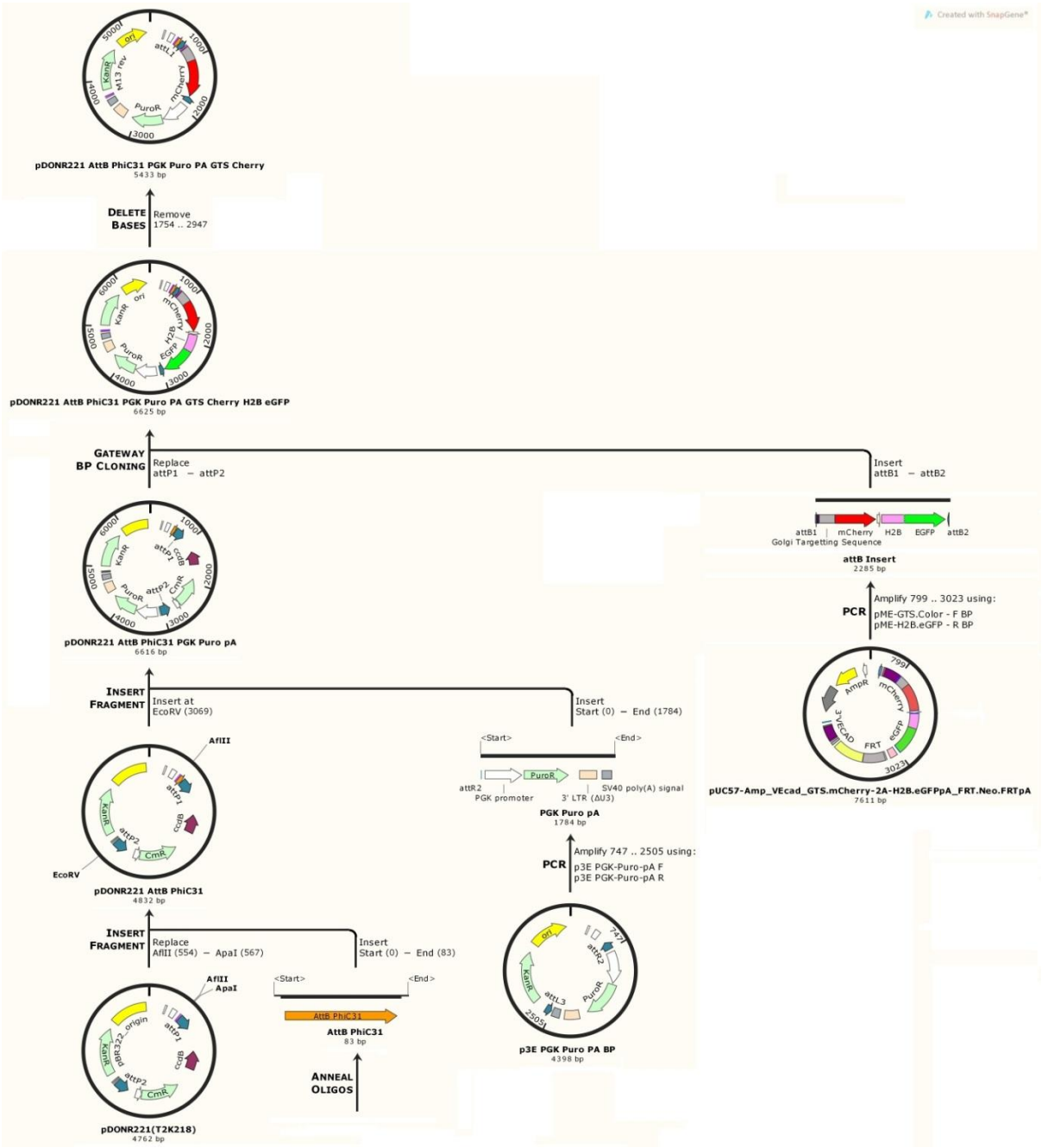



**Figure A5.** Schematic overview of the cloning process that resulted in the generation of the pDONR221 AttB PhiC31 PGK Puro PA GTS Cherry plasmid. The figure shows the different vectors used and the strategy followed.

# Plasmid Maps and Sequences

## p3E Neo PolyA

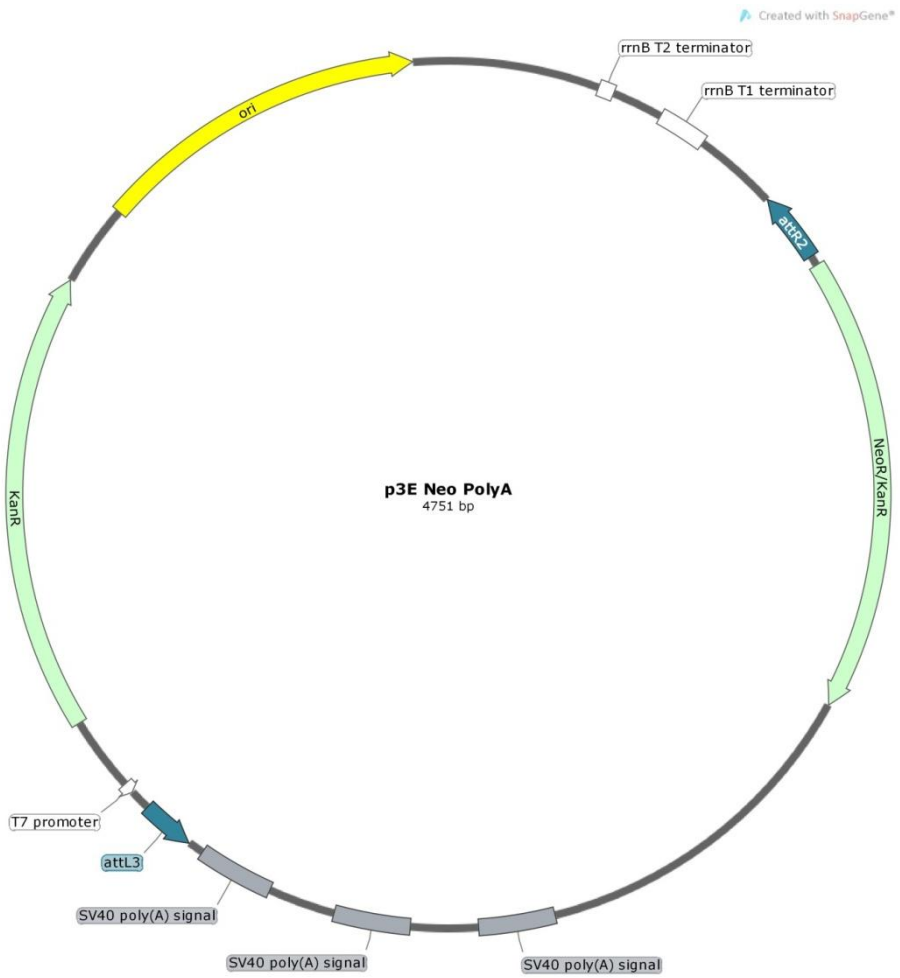



**Figure A6.** p3E Neo PolyA vector map. Features were automatically annotated using SnapGene® Software.

```
CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCTGCAGCTCTAGAGCTCGAATTCTACAGGTCAC
TAATACCATCTAAGTAGTTGGTTCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGAT
ATTTATATCATTTTACGTTTCTCGTTCAACTTTCTTGTACAAAGTGGCTTTCGACCTGCAGCCAATATGGGATCGGCCATTGAACAAGATGGATTGCACG
CAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCA
GGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTT
CCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTC
CTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGA
GCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAG
GCGCGCATGCCCGACGGCGATGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACT
GTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGT
GCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGGATCCGCTGTAAGTCTGCAGAAATTGA
TGATCTATTAAACAATAAAGATGTCCACTAAAATGGAAGTTTTTCCTGTCATACTTTGTTAAGAAGGGTGAGAACAGAGTACCTACATTTTGAATGGAAG
GATTGGAGCTACGGGGGTGGGGGTGGGGTGGGATTAGATAAATGCCTGCTCTTTACTGAAGGCTCTTTACTATTGCTTTATGATAATGTTTCATAGTTGG
ATATCATAATTTAAACAAGCAAAACCAAATTAAGGGCCAGCTCATTCCTCCCACTCATGATCTATAGATCTATAGATCTCTCGTGGGATCATTGTTTTTC
TCTTGATTCCCACTTTGTGGTTCTAAGTACTGTGGTTTCCAAATGTGTCAGTTTCATAGCCTGAAGAACGAGATCAGCAGCCTCTGTTCCACATACACTT
CATTCTCAGTATTGTTTTGCCAAGTTCTAATTCCATCAGAAGCTTGCAGATCTGCGACTCTAGAGGATCTGCGACTCTAGAGGATCATAATCAGCCATAC
CACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATT
GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA
```


ATGTATCTTATCATGTCTGGATCTGCGACTCTAGAGGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTC
CCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCA
CAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTGCGACTCTAGAGGATCATAATC
AGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACT
TGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAA
ACTCATCAATGTATCTTATCATGTCTGGATCCCCATCAAGCTGATCCGGAACCCTTAATCAACTTTATTATACAAAGTTGGCATTATAAAAAAGCATTGC
TTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGGAGCTCCATGGTAGCGTTAACGCGGCCGCGATATCCCCTATAGT
GAGTCGTATTACATGGTCATAGCTGTTTCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAA
CAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGC
TGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTG
AAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTA
TCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGT
TGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCA
CGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCAT
TCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGG
AATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGAT
AATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGA
CTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTC
TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTT
TTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGC
CTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAA
AGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCT
GGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAA
CGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT ••• 4751

## p3E PGK Puro PA BP

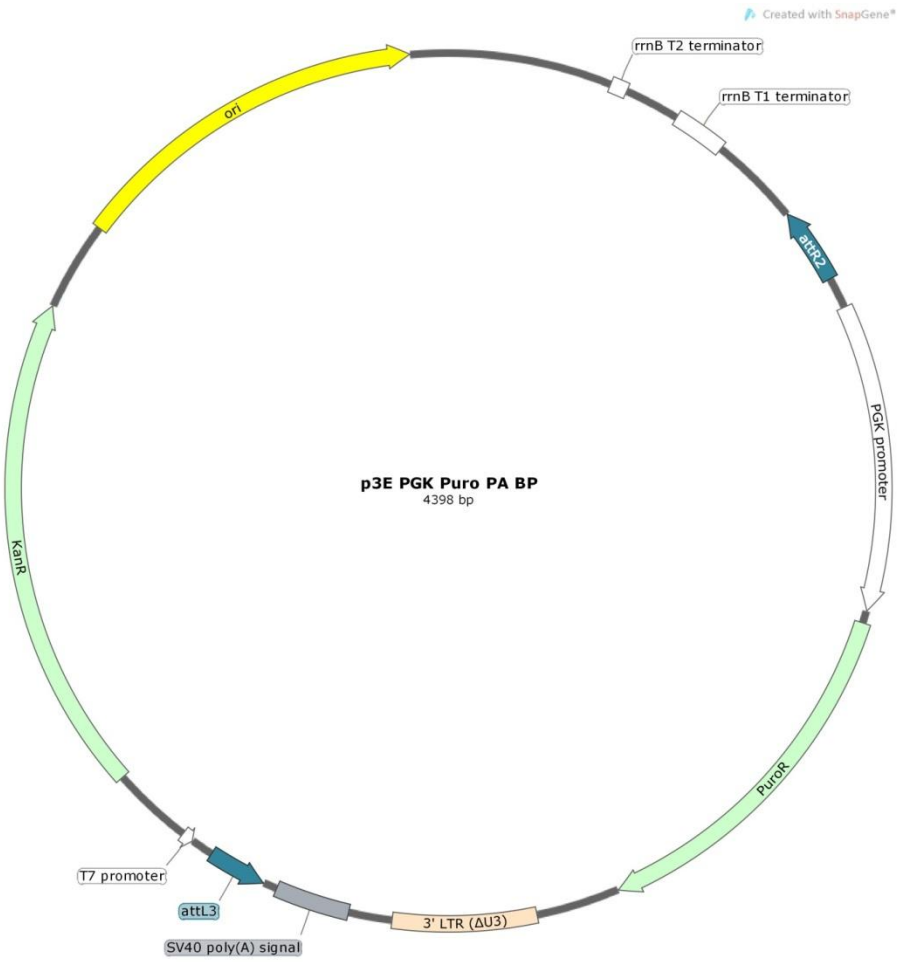



**Figure A7.** p3E PGK Puro PA BP vector map. Features were automatically annotated using SnapGene® Software.

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCTGCAGCTCTAGAGCTCGAATTCTACAGGTCAC
TAATACCATCTAAGTAGTTGGTTCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGAT
ATTTATATCATTTTACGTTTCTCGTTCAACTTTCTTGTACAAAGTGGCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGGAATTCTACCG
GGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACAT
TCCACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGC
AGCTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGATGGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCTTTG
GGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGG
GCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCGCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGA
CCTGCAGCCCAAGCTTACCATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCGCCGCCGCGTTC
GCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGCTCGACA
TCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCG
CATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACC
GTCGGCGTCTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCCTGG
AGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGAC
CCGCAAGCCCGGTGCCTGACGCCCGCCCCACGACCCGCAGCGCCCGACCGAAAGGAGCGCACGACCCCATGCATCTCGAGGGCCCGGTACCTTTAAGACC
AATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAGCTCACTCCCAACGAAGACAAGATCTGCTTTTT
GCTTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAG
TGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTAGTAGTTC
ATGTCATCTTATTATTCAGTATTTATAACTTGCAAAGAAATGAATATCAGAGAGTGAGAGGAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGC
AATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGCTCTAGC
TATCCCCAACTTTATTATACAAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTAT
TTGGAGCTCCATGGTAGCGTTAACGCGGCCGCGATATCCCCTATAGTGAGTCGTATTACATGGTCATAGCTGTTTCCTGGCAGCTCTGGCCCGTGTCTCA


AAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCAT
ATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGA
CAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACT
AAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACA
GCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATT
GTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGG
CTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATT
TTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGT
TTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTA
ATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCG
TCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCAC
CGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCT
AGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGC
GAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAG
GGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTT
TTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT

## p5E-MCS (T2K228)

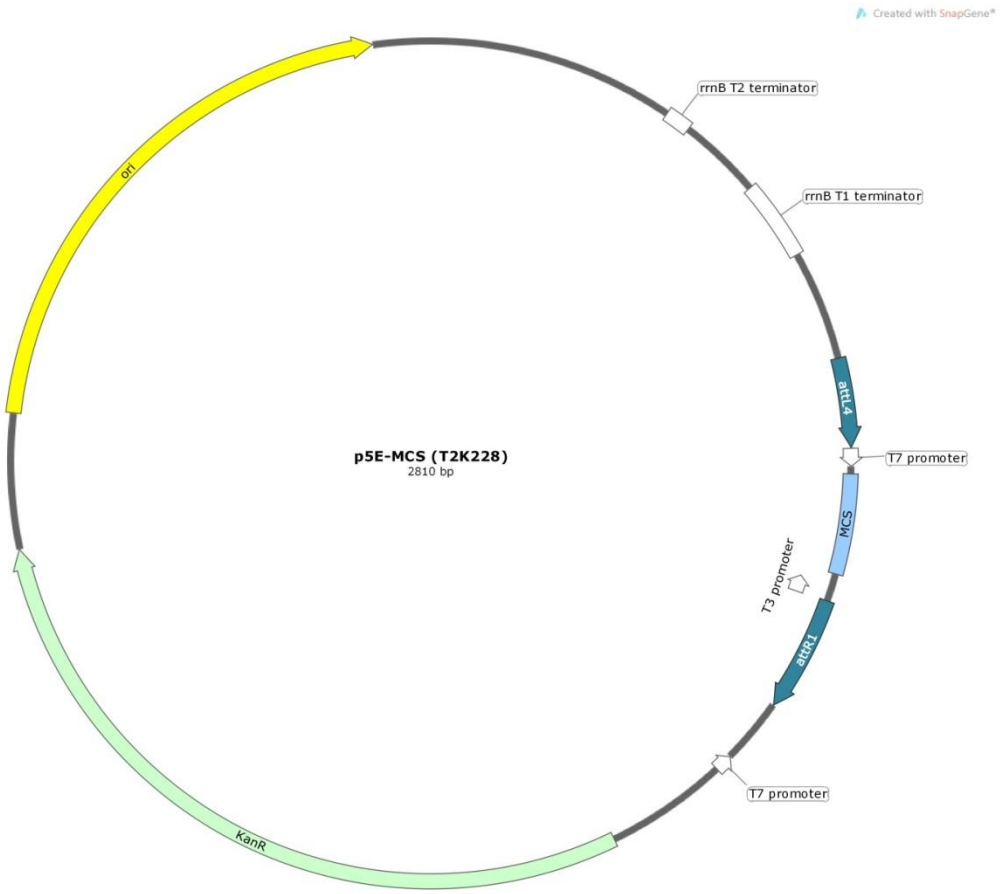



**Figure A8.** p5E-MCS (T2K228) vector map. Features were automatically annotated using SnapGene® Software.

```
CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGCGTTAACGCTACCATGGAGCTCCAAATAATG
ATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTATAGAAAAGTTGGTAATACGACTC
ACTATAGGGCGAATTGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGA
GCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTCAAGTTTGTACAAAAAAGTTGAACGAGAAACGTAAAATGATATAA
ATATCAATATATTAAATTAGATTTTGCATAAAAAACAGACTACATAATACTGTAAAACACAACATATGCAGTCACTATGAACCAACTACTTAGATGGTAT
TAGTGACCTGTAGAATTCGAGCTCTAGAGCTGCAGGGCGGCCGCGATATCCCCTATAGTGAGTCGTATTACATGGTCATAGCTGTTTCCTGGCAGCTCTG
GCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGT
GTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGC
AATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGA
GATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATC
CCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTC
CTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGA
CGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTT
GATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACT
GCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGA
TGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAAC
GTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA
CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAA
TACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC
CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTT
GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG
CTCACATGTT ••• 2810
```

## p5E MCS CAGGS AttPhiC31

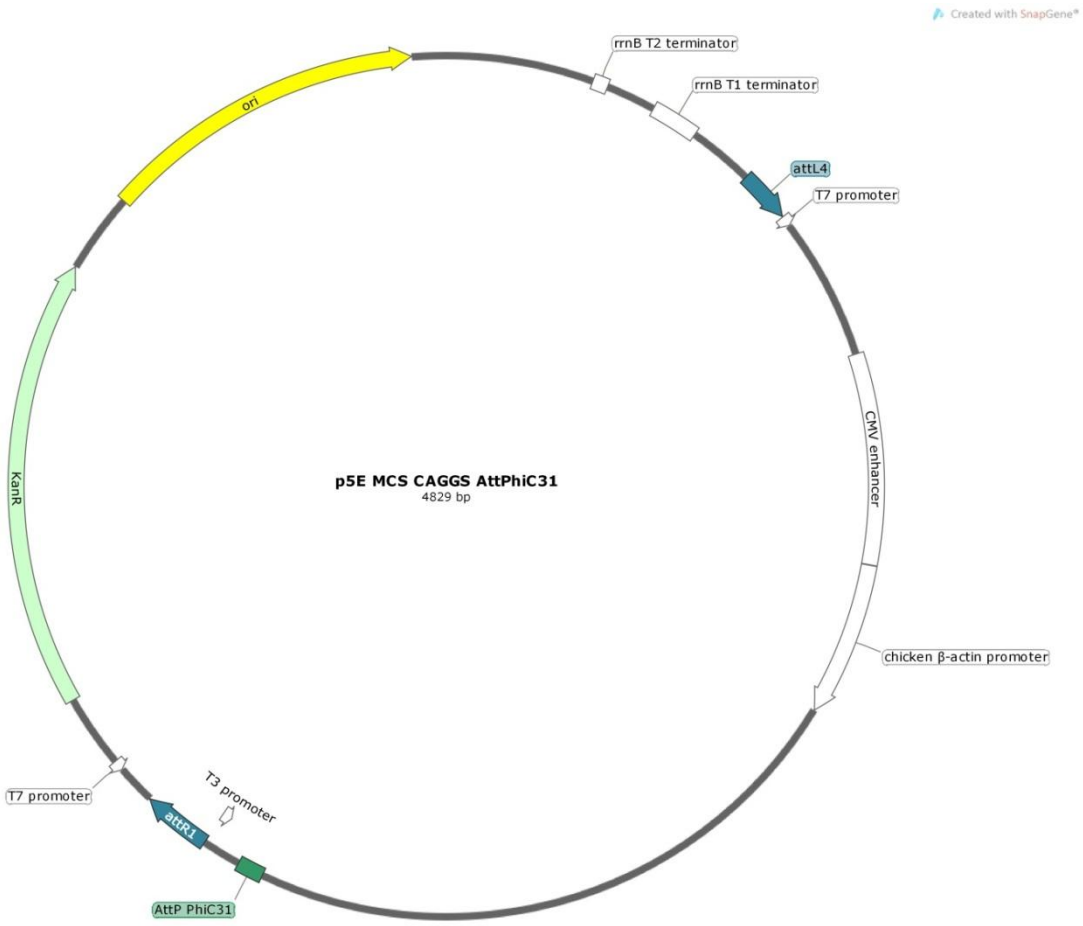
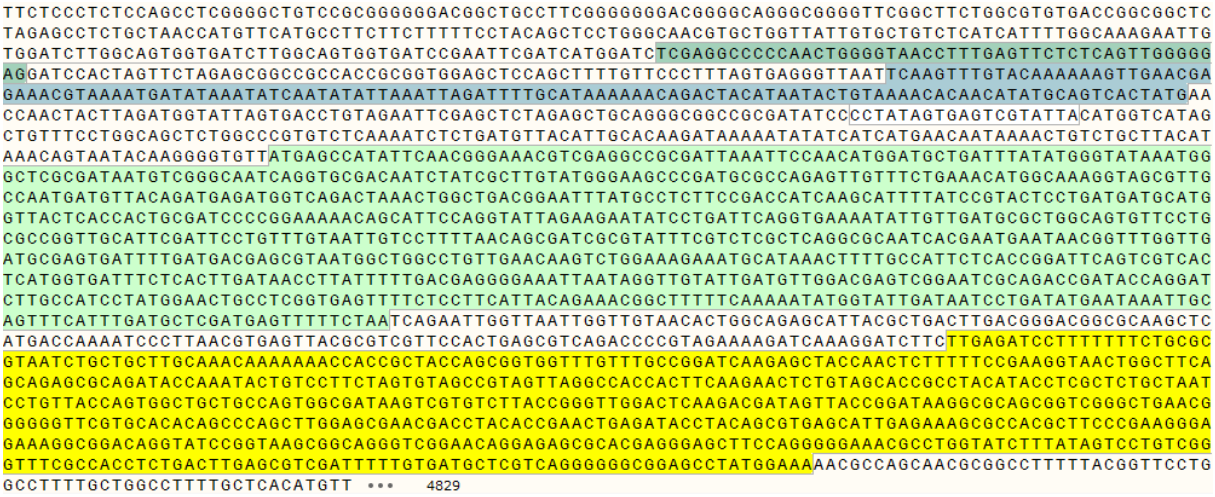



**Figure A9.** p5E MCS CAGGS AttP PhiC31 vector map. Features were automatically annotated using SnapGene® Software.

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGCGTTAACGCTACCATGGAGCTCCAAATAATG
ATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTATAGAAAAGTTGGTAATACGACTC
ACTATAGGGCGAATTGGGTACCGGGCCCCCCCTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCAT
AGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTG
CATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAA
TCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCC
CATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGAC
TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCTCCC
CACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGA
GGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCCC
TATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGAC
TGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTTGTTTCTTTTCTGTGGCT
GCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGC
TCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGC
GGTGCGGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCC
CCCTGCACCCCCCTTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGT
GGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCGAGG
CGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGCGC
CGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCC


TTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTC
TAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTG
TGGATCTTGGCAGTGGTGATCTTGGCAGTGGTGATCCGAATTCGATCATGGATCTCGAGGCCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG
AGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTCAAGTTTGTACAAAAAAGTTGAACGA
GAAACGTAAAATGATATAAATATCAATATATTAAATTAGATTTTGCATAAAAAACAGACTACATAATACTGTAAAACACAACATATGCAGTCACTATGAA
CCAACTACTTAGATGGTATTAGTGACCTGTAGAATTCGAGCTCTAGAGCTGCAGGGCGGCCGCGATATCCCCTATAGTGAGTCGTATTACATGGTCATAG
CTGTTTCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACAT
AAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGG
GCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTG
CCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATG
GTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTG
CGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTG
ATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCAC
TCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGAT
CTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGC
AGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTC
ATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCA
GCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAAT
CCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGA
GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTG
GCCTTTTGCTGGCCTTTTGCTCACATGTT ••• 4829

## pCAGGS-m2G

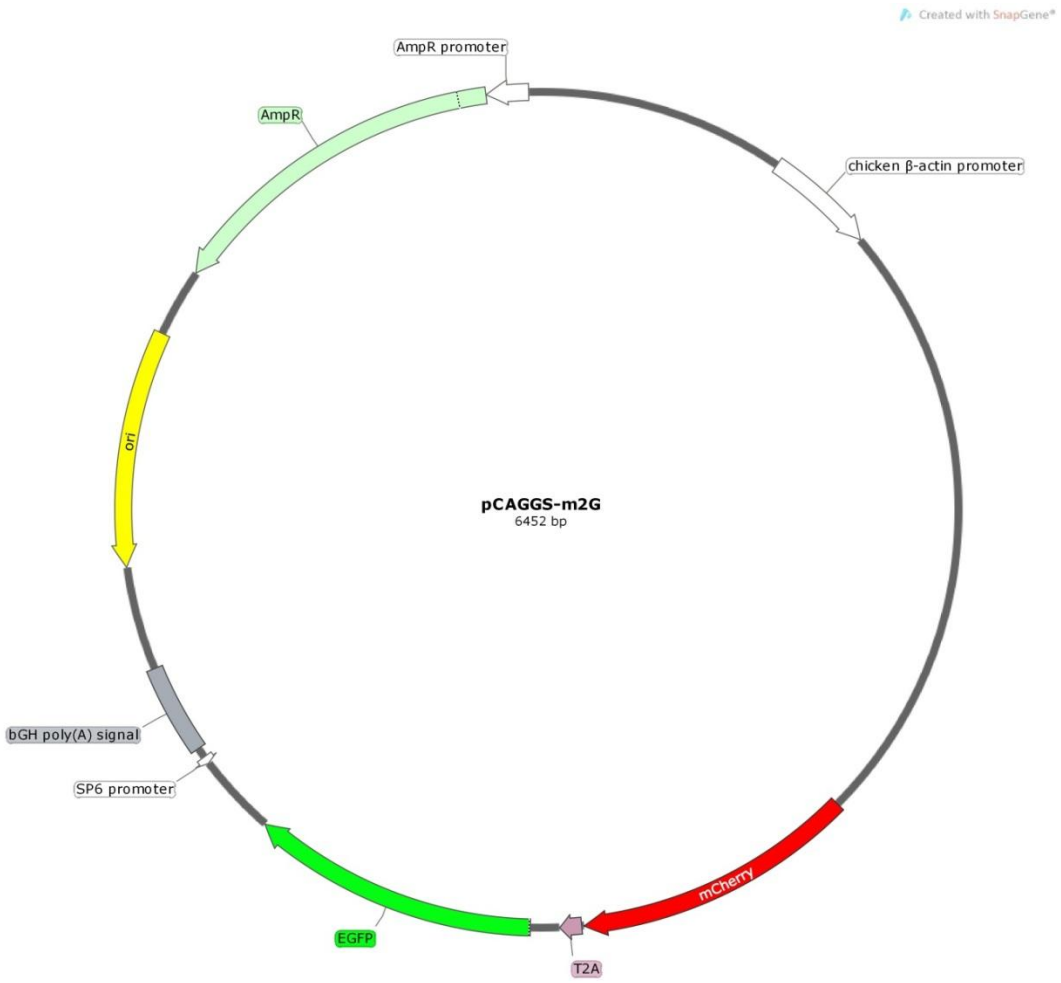



**Figure A10.** pCAGGS-m2G vector map. Features were automatically annotated using SnapGene® Software.

CGACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT
TGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGC
GCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATAT
ATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCATTTACCATGGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAA
TTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCG
GCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCG
CTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGG
CCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTAAAGGGCTCCGGGAGGGCCCTTTG
TGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCG
GCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGC
GGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGG
CTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGGAGGTGGGGGTGCCGCGCGGGGCGGGGCCGCCTCG
GGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCG
AGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGC
GCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCT
GCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCT
ACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTGTGGATCTTGGCAGTGGTGATCTTGGCAGTGGTGATCCGAATT
CGATCATGGATCTCGAGCCACCATGCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGC
GGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCA
TGGGCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGTGAGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCAG
GGAGATCCAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCGTCACCAAGTACACCAGCGCTAAG


6452

## pCS2P+PhiC31o

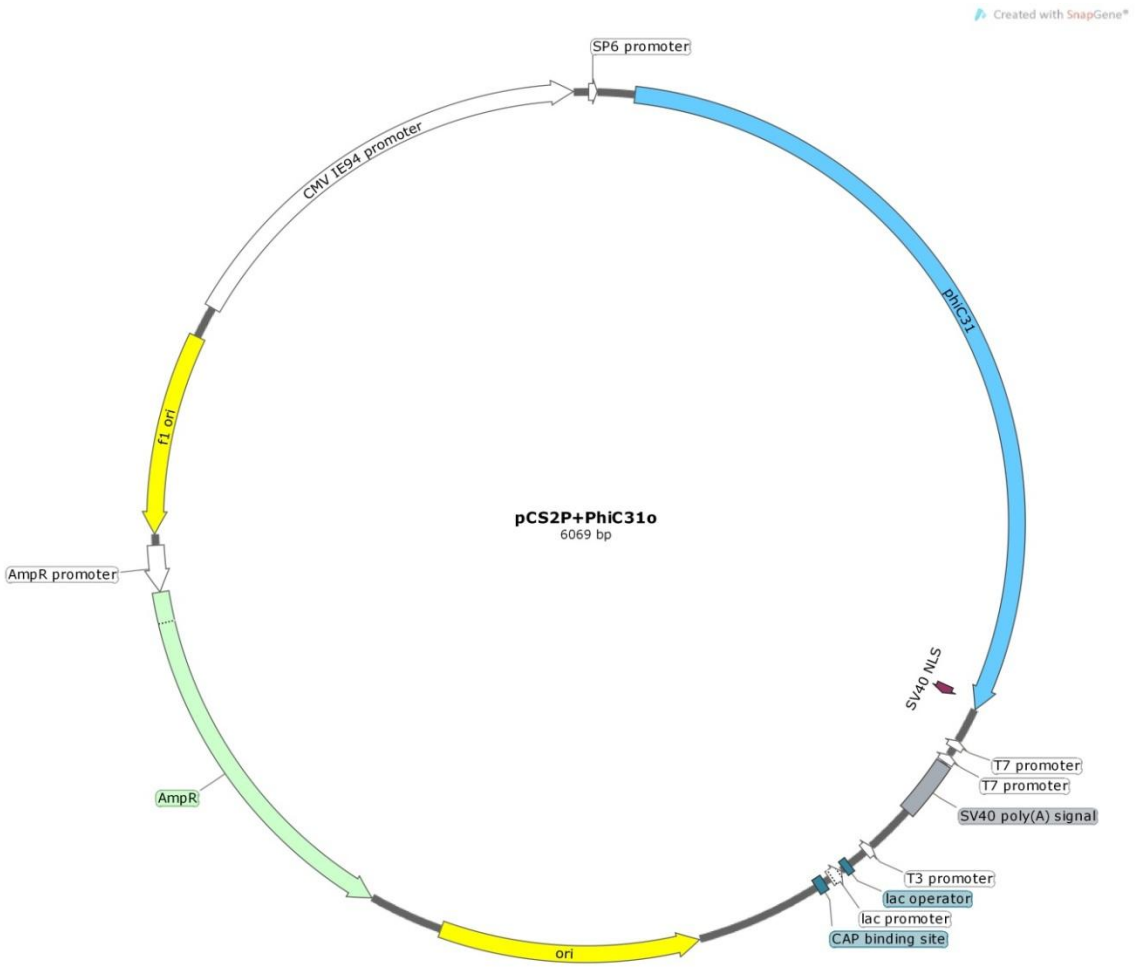
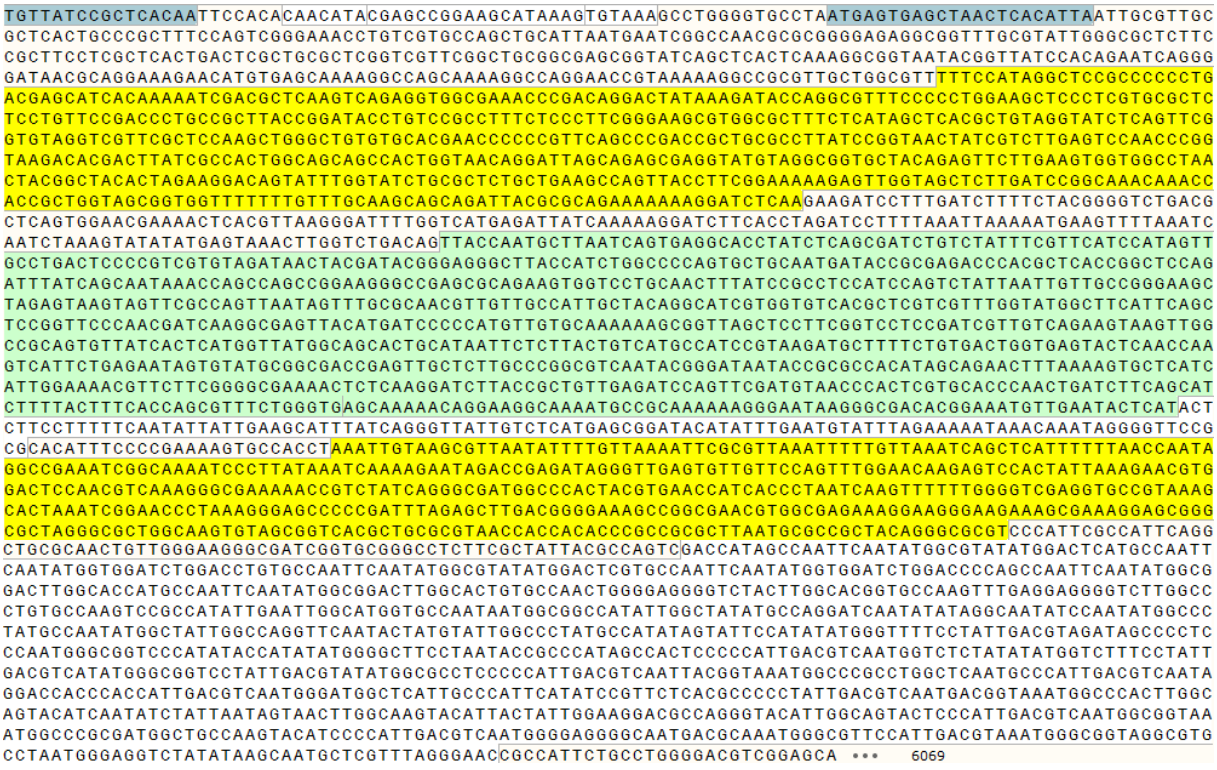



**Figure A11.** pCS2P+PhiC31o vector map. Features were automatically annotated using SnapGene® Software.

AGCTTGATTTAGGTGACACTATAGAATACAAGCTACTTGTTCTTTTTGCAGGATCCCATCGATTCGAATTCGAGCTCGGTACCCGACACGGTACCGATAT
CGCCGCCACCATGGATACCTACGCCGGAGCCTACGACAGACAGAGCCGGGAGAGAGAGAACAGCAGCGCCGCCAGCCCCGCCACCCAGAGAAGCGCCAAC
GAGGATAAGGCCGCCGATCTGCAGAGAGAGGTGGAGAGGGACGGCGGCAGATTCAGATTTGTGGGCCACTTCAGCGAGGCCCCTGGCACCAGCGCCTTCG
GCACCGCCGAGAGACCCGAGTTCGAGAGAATCCTGAACGAGTGTAGGGCCGGCAGGCTGAACATGATCATCGTGTACGACGTGTCCCGGTTCAGCAGGCT
GAAGGTGATGGACGCCATCCCTATCGTGTCCGAGCTGCTGGCCCTGGGCGTGACCATCGTGTCCACCCAGGAAGGCGTCTTTAGACAGGGCAACGTGATG
GACCTGATCCACCTGATCATGAGGCTGGACGCCAGCCACAAGGAGAGCAGCCTGAAGAGCGCCAAGATCCTGGACACCAAGAACCTGCAGAGGGAGCTGG
GCGGCTATGTGGGCGGCAAGGCCCCCTACGGCTTCGAGCTGGTGTCCGAGACCAAGGAGATCACCCGGAACGGCAGGATGGTGAACGTGGTGATCAACAA
GCTGGCCCACAGCACCACCCCCCTGACCGGCCCCTTCGAGTTTGAGCCCGACGTGATCAGGTGGTGGTGGCGGGAGATCAAGACCCACAAGCACCTGCCT
TTCAAGCCCGGCAGCCAGGCCGCCATCCACCCCGGCAGCATCACCGGCCTGTGTAAGAGAATGGACGCCGACGCCGTGCCCACCAGAGGCGAGACCATCG
GCAAGAAAACCGCCAGCAGCGCCTGGGACCCCGCCACCGTGATGAGAATCCTGAGGGACCCTAGGATCGCCGGCTTCGCCGCCGAGGTGATCTACAAGAA
GAAGCCCGACGGCACCCCCACCACCAAGATCGAGGGCTACAGAATCCAGAGAGACCCCATCACCCTGAGACCTGTGGAGCTGGACTGTGGCCCTATCATC
GAGCCTGCCGAGTGGTACGAGCTGCAGGCCTGGCTGGACGGCAGAGGCAGAGGCAAGGGCCTGAGCAGAGGCCAGGCCATCCTGAGCGCCATGGACAAGC
TGTACTGTGAGTGTGGCGCCGTGATGACCAGCAAGAGAGGCGAGGAGAGCATCAAGGACAGCTACCGGTGCCGGAGAAGAAAGGTGGTGGACCCCAGCGC
CCCTGGCCAGCACGAGGGCACCTGTAATGTGAGCATGGCCGCCCTGGACAAGTTCGTGGCCGAGCGGATCTTCAACAAGATCCGGCACGCCGAGGGCGAC
GAGGAGACCCTGGCCCTGCTGTGGGAGGCCGCCAGAAGATTCGGCAAGCTGACCGAGGCCCCCGAGAAGAGCGGCGAGAGGGCCAACCTGGTGGCCGAGA
GAGCCGACGCCCTGAACGCCCTGGAGGAGCTGTACGAGGACAGAGCCGCCGGAGCCTATGACGGCCCTGTGGGCAGGAAGCACTTCAGAAAGCAGCAGGC
CGCCCTGACCCTGAGACAGCAGGGCGCCGAGGAAAGACTGGCCGAGCTGGAGGCCGCCGAGGCCCCTAAGCTGCCCCTGGATCAGTGGTTCCCCGAGGAT
GCCGACGCCGACCCCACCGGCCCCAAGTCCTGGTGGGGCAGAGCCAGCGTGGACGACAAGAGGGTGTTCGTGGGCCTGTTCGTGGATAAGATCGTGGTGA
CCAAGAGCACCACCGGCAGGGGCCAGGGCACCCCCATCGAGAAGAGAGCCAGCATCACCTGGGCCAAGCCTCCCACCGACGACGACGAGGATGACGCCCA
GGACGGCACCGAGGACGTGGCCGCCCCTAAGAAAAAGCGGAAAGTGTGATAGATATCGAGCTCGTCTCGGGGATCCTCTAGAGTCGAGGGGCCCAAGCTT
ACGCGTGCATGCGACGTCATAGCTCTCTCCCTATAGTGAGTCGTATTATAAGCTAGATTCTGCAGCCCTATAGTGAGTCGTATTACGTAGATCCAGACAT
GATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACC
ATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAATTCGCGGCCGCGG
CGCCAATGCATTGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAAT


TGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGC
GCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTC
CGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGG
GATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTCCATAGGCTCCGCCCCCCTG
ACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTC
TCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCG
GTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACC
ACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTT
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAG
ATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGC
TAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGG
CCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAA
GTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATC
ATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT
CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACT
CTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCG
CGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATA
GGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTG
GACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAG
CACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGG
CGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGG
CTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGTCGACCATAGCCAATTCAATATGGCGTATATGGACTCATGCCAATT
CAATATGGTGGATCTGGACCTGTGCCAATTCAATATGGCGTATATGGACTCGTGCCAATTCAATATGGTGGATCTGGACCCCAGCCAATTCAATATGGCG
GACTTGGCACCATGCCAATTCAATATGGCGGACTTGGCACTGTGCCAACTGGGGAGGGGTCTACTTGGCACGGTGCCAAGTTTGAGGAGGGGTCTTGGCC
CTGTGCCAAGTCCGCCATATTGAATTGGCATGGTGCCAATAATGGCGGCCATATTGGCTATATGCCAGGATCAATATATAGGCAATATCCAATATGGCCC
TATGCCAATATGGCTATTGGCCAGGTTCAATACTATGTATTGGCCCTATGCCATATAGTATTCCATATATGGGTTTTCCTATTGACGTAGATAGCCCCTC
CCAATGGGCGGTCCCATATACCATATATGGGGCTTCCTAATACCGCCCATAGCCACTCCCCCATTGACGTCAATGGTCTCTATATATGGTCTTTCCTATT
GACGTCATATGGGCGGTCCTATTGACGTATATGGCGCCTCCCCCATTGACGTCAATTACGGTAAATGGCCCGCCTGGCTCAATGCCCATTGACGTCAATA
GGACCACCCACCATTGACGTCAATGGGATGGCTCATTGCCCATTCATATCCGTTCTCACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCACTTGGC
AGTACATCAATATCTATTAATAGTAACTTGGCAAGTACATTACTATTGGAAGGACGCCAGGGTACATTGGCAGTACTCCCATTGACGTCAATGGCGGTAA
ATGGCCCGCGATGGCTGCCAAGTACATCCCCATTGACGTCAATGGGGAGGGGCAATGACGCAAATGGGCGTTCCATTGACGTAAATGGGCGGTAGGCGTG
CCTAATGGGAGGTCTATATAAGCAATGCTCGTTTAGGGAACCGCCATTCTGCCTGGGGACGTCGGAGCA ••• 6069

## pDONR221 (T2K218)

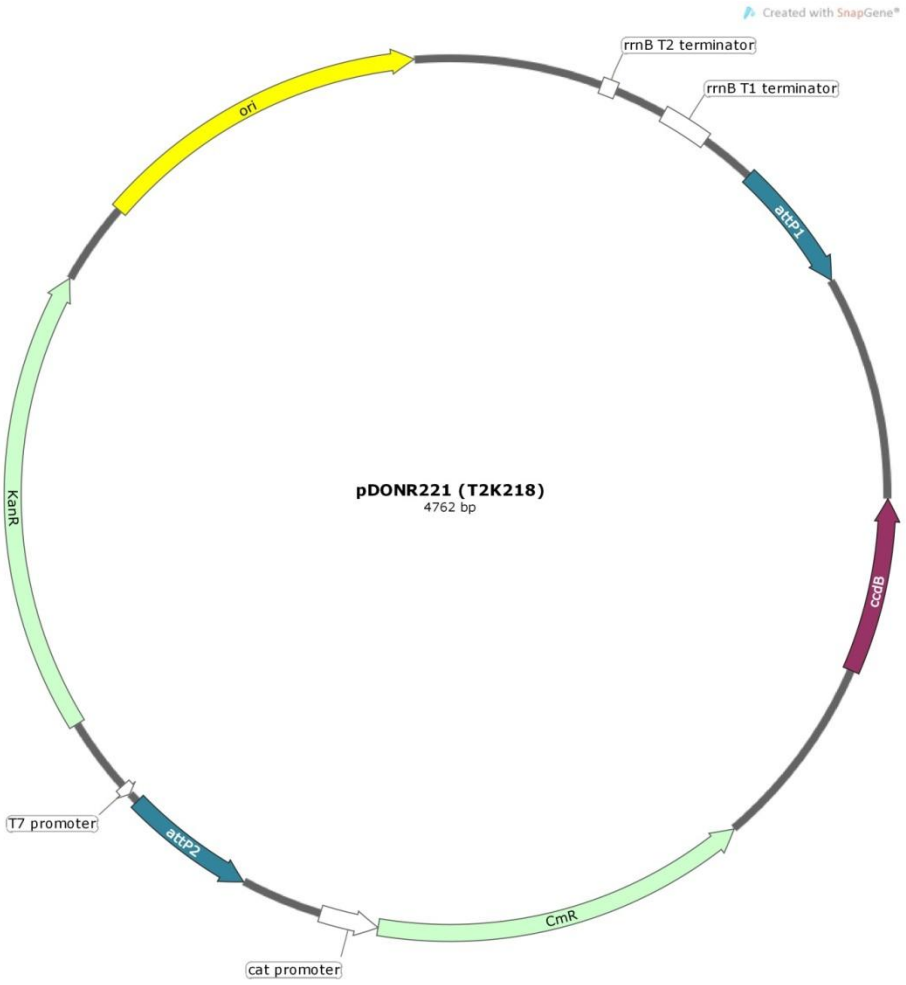



**Figure A12.** pDONR221 (T2K218) vector map. Features were automatically annotated using SnapGene® Software.

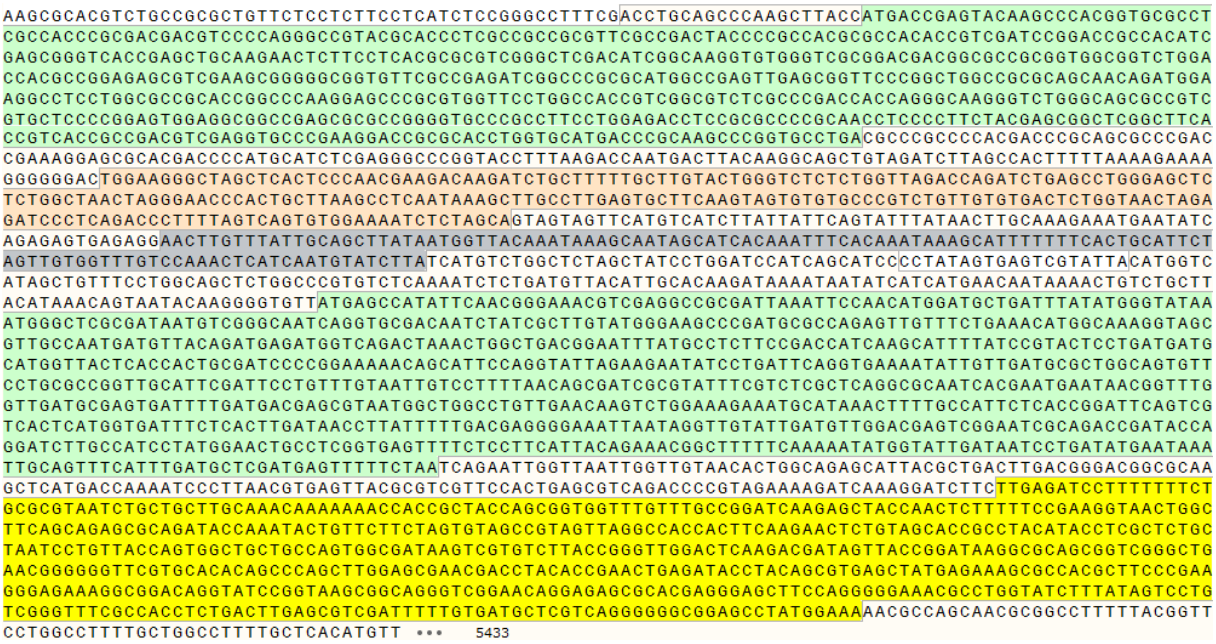
CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGTGAC
CTGTTCGTTGCAACACATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCTGAACGAGAAACGTAAAATGATATAAATATCAATATA
TTAAATTAGATTTTGCATAAAAAACAGACTACATAATACTGTAAAACACAACATATCCAGTCACTATGAATCAACTACTTAGATGGTATTAGTGACCTGT
AGTCGACCGACAGCCTTCCAAATGTTCTTCGGGTGATGCTGCCAACTTAGTCGACCGACAGCCTTCCAAATGTTCTTCTCAAACGGAATCGTCGTATCCA
GCCTACTCGCTATTGTCCTCAATGCCGTATTAAATCATAAAAAGAAATAAGAAAAAGAGGTGCGAGCCTCTTTTTTGTGTGACAAAATAAAAACATCTAC
CTATTCATATACGCTAGTGTCATAGTCCTGAAAATCATCTGCATCAAGAACAATTTCACAACTCTTATACTTTTCTCTTACAAGTCGTTCGGCTTCATCT
GGATTTTCAGCCTCTATACTTACTAAACGTGATAAAGTTTCTGTAATTTCTACTGTATCGACCTGCAGACTGGCTGTGTATAAGGGAGCCTGACATTTAT
ATTCCCCAGAACATCAGGTTAATGGCGTTTTTGATGTCATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAGACCGGCACACTGGCCA
TATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGG
GATCACCATCCGTCGCCCGGGCGTGTCAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACCTTAAACTGC
ATTTCACCAGCCCCTGTTCTCGTCAGCAAAAGAGCCGTTCATTTCAATAAACCGGGCGACCTCAGCCATCCCTTCCTGATTTTCCGCTTTCCAGCGTTCG
GCACGCAGACGACGGGCTTCATTCTGCATGGTTGTGCTTACCAGACCGGAGATATTGACATCATATATGCCTTGAGCAACTGATAGCTGTCGCTGTCAAC
TGTCACTGTAATACGCTGCTTCATAGCATACCTCTTTTTGACATACTTCGGGTATACATATCAGTATATATTCTTATACCGCAAAAATCAGCGCGCAAAT
ACGCATACTGTTATCTGGCTTTTAGTAAGCCGGATCCACGCGGCGTTTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCT
GCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACG
GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTT
TAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGA
AAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGA
GCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGG


TCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTT
CTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACG
TGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCA
CAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTAGTCGACTACAGGTCACTAATACCATCTAAGTAGTTGATTCATAGTGACTGGAT
ATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTCTTG
TACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGCCATCCAGCTGATA
TCCCCTATAGTGAGTCGTATTACATGGTCATAGCTGTTTCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAATAAT
ATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCC
AACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAG
AGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCAT
CAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGT
GAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTC
AGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAA
ACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTT
GGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAAT
ATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAG
CATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATC
AAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGC
TACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTC
TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAG
TTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAA
AACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT ••• 4762

## pDONR221 AttB PhiC31 PGK Puro PA GTS Cherry


**Figure A13.** pDONR221 AttB PhiC31 PGK Puro PA GTS Cherry vector map. Features were automatically annotated using SnapGene® Software.

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGTACTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGG
CTCCCCGGGCGCGTACTCCACCTCACCCATCTCGGCCCCAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATG
CTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTTACCAGATCAGCTCCTCCACGAAGGGCTAGCGCCGCCAGCATGAGGCTTCGGGAGCCGCTCC
TGAGCGGCAGCGCCGCGATGCCAGGCGCGTCCCTACAGCGGGCCTGCCGCCTGCTCGTGGCCGTCTGCGCTCTGCACCTTGGCGTCACCCTCGTTTACTA
CCTGGCTGGCCGCGACCTGAGCCGCCTGCCCCAACTGGTCGGAGTCTCCACACCGCTGCAGGGCGGCTCGAACAGTGCCGCCGCCATCGGGCAGTCCTCC
GGGGAGCTCCGGACCGGAGGGGCCAAGGATCCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCT
TCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAA
GGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC
GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGC
AGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTC
CGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCAC
TACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGT
ACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCACGCGTATGTTCACCCTAACCCAGCTTTCTTGTACAAAGTTGGCA
TTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGCCATCCAGCTGATCGATCTGGATCCGC
TCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGGAATTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGC
AGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCC
TTCGCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTA
GTCTCGTGCAGATGGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGAG
GCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAA


AAGCGCACGTCTGCCGCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCAGCCCAAGCTTACCATGACCGAGTACAAGCCCACGGTGCGCCT
CGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCACATC
GAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTCTGGA
CCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGA
AGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTC
GTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCA
CCGTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGACGCCCGCCCACGACCCGCAGCGCCCGAC
CGAAAGGAGCGCACGACCCCATGCATCTCGAGGGCCCGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAA
GGGGGGACTGGAAGGGCTAGCTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTC
TCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGA
GATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTAGTAGTTCATGTCATCTTATTATTCAGTATTTATAACTTGCAAAGAAATGAATATC
AGAGAGTGAGAGGAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCT
AGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGCTCTAGCTATCCTGGATCCATCAGCATCCCCTATAGTGAGTCGTATTACATGGTC
ATAGCTGTTTCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAATAATATCATCATGAACAATAAAACTGTCTGCTT
ACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAA
ATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGC
GTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATG
CATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTT
CCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTG
GTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCG
TCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCA
GGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAA
TTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAA
GCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCT
GCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGC
TTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC
TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTG
AACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAA
GGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTG
TCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTT
CCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT ••• 5433

## pDONRP2R-P3 (T2K220)

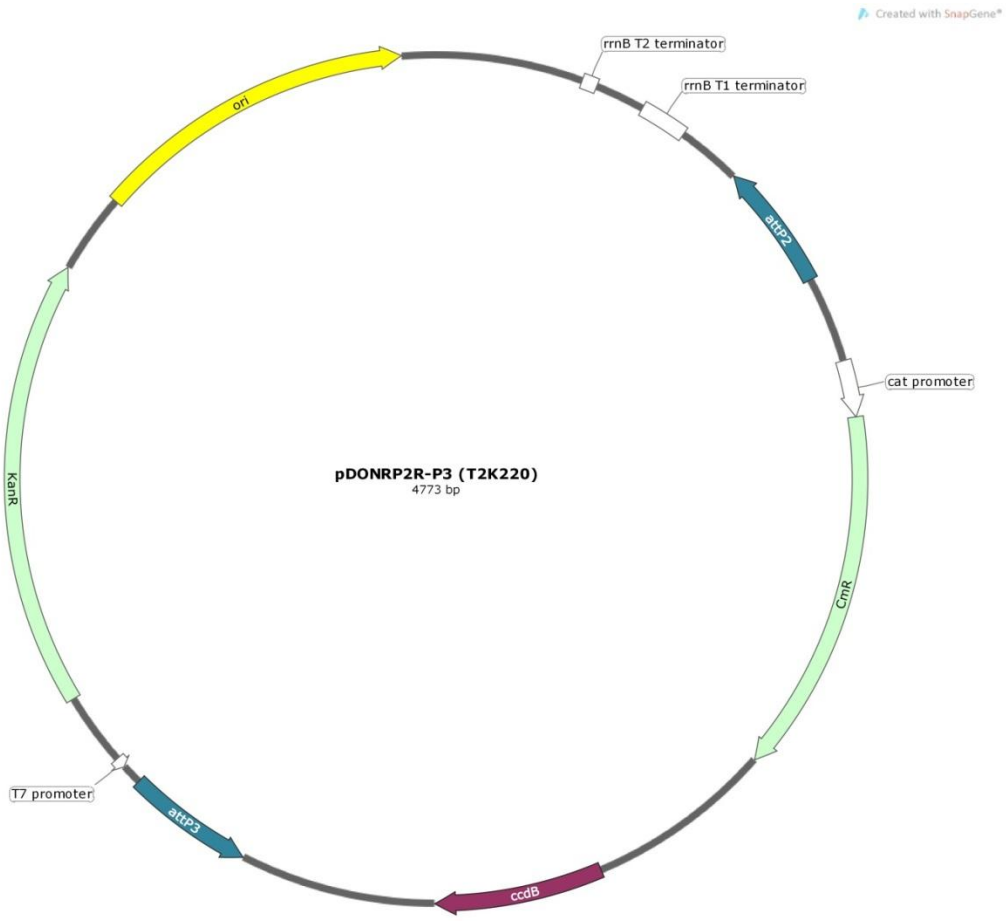



**Figure A14.** pDONRP2R-P3 (T2K220) vector map. Features were automatically annotated using SnapGene® Software.

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCTGCAGCTCTAGAGCTCGAATTCTACAGGTCAC
TAATACCATCTAAGTAGTTGATTCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGAT
ATTTATATCATTTTACGTTTCTCGTTCAACTTTCTTGTACAAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATC
AGTCAAAATAAAATCATTATTTGGAGCTCTAGAGCGTCGACTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATC
ACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTC
CAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACT
GGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGG
ATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATT
CCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGG
AGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTG
AGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCAT
GGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAAT
GAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATCGCGTGGATCCGGCTTACTAAAAGCCAGATAACAGTATGCGTATTTGCGCGCTGATT
TTTGCGGTATAAGAATATATACTGATATGTATACCCGAAGTATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCGACAGC
TATCAGTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGTCTGCGTGCCGAACGCTGGAAAGC
GGAAAATCAGGAAGGGATGGCTGAGGTCGCCCGGTTTATTGAAATGAACGGCTCTTTTGCTGACGAGAACAGGGACTGGTGAAATGCAGTTTAAGGTTTA
CACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCA
CGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGG
TCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGGCTC
CCTTATACACAGCCAGTCTGCAGGTCGATACAGTAGAAATTACAGAAACTTTATCACGTTTAGTAAGTATAGAGGCTGAAAATCCAGATGAAGCCGAACG
ACTTGTAAGAGAAAAGTATAAGAGTTGTGAAATTGTTCTTGATGCAGATGATTTTCAGGACTATGACACTAGCGTATATGAATAGGTAGATGTTTTTATT
TTGTCACACAAAAAAGAGGCTCGCACCTCTTTTTCTTATTTCTTTTTATGATTTAATACGGCATTGAGGACAATAGCGAGTAGGCTGGATACGACGATTC


CGTTTGAGAAGAACATTTGGAAGGCTGTCGGTCGAGCTCGAATTCTACAGGTCACTAATACCATCTAAGTAGTTGATTCATAGTGACTGCATATGTTGTG
TTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAACTTTATTATACAAAGT
TGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGGAGCTCCATGGTAGCGTTAACG
CGGCCGCGATATCCCCTATAGTGAGTCGTATTACATGGTCATAGCTGTTTCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAA
GATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGC
GATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCC
CGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCT
CTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATC
CTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATT
TCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAA
GAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTT
GTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCT
TTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAA
CACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCG
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACT
CAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA
CCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG
GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT ••• 4773

## pME PhiC31 2A

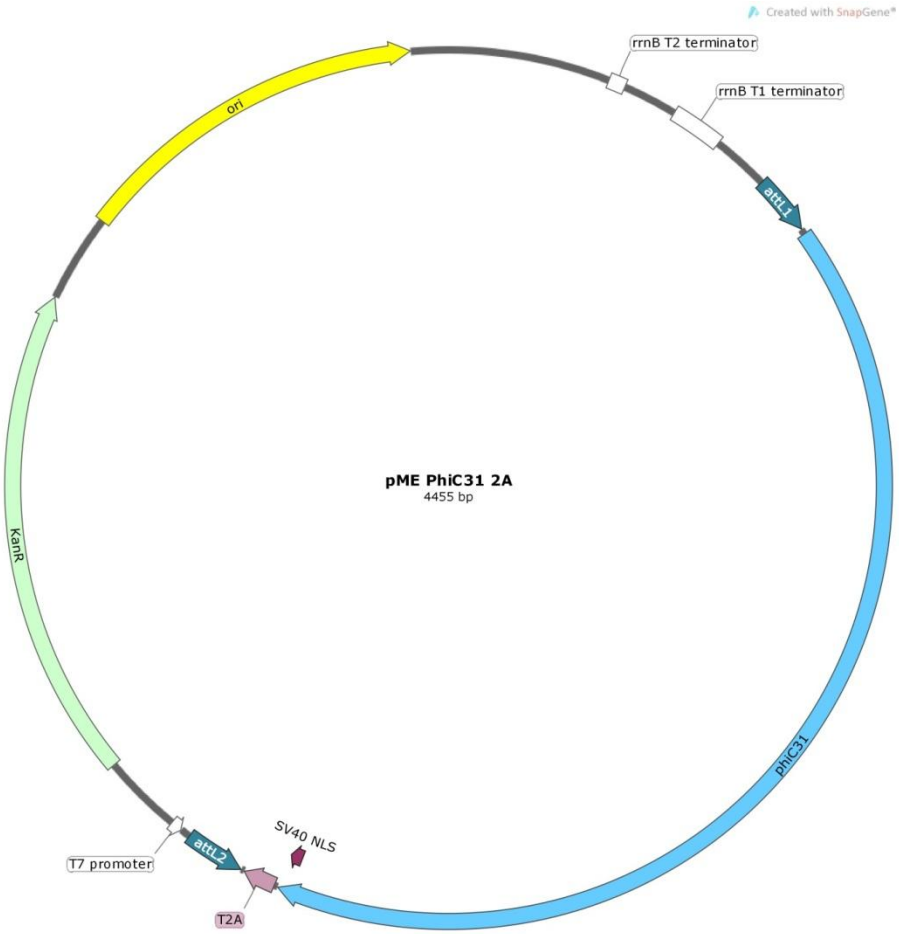



**Figure A15.** pME PhiC31 2A vector map. Features were automatically annotated using SnapGene® Software.

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTA
GTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGT
TAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGTGAC
CTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTGCCGCCACCATGGATACCTACGCCGGAGCCTA
CGACAGACAGAGCCGGGAGAGAGAGAACAGCAGCGCCGCCAGCCCCGCCACCCAGAGAAGCGCCAACGAGGATAAGGCCGCCGATCTGCAGAGAGAGGTG
GAGAGGGACGGCGGCAGATTCAGATTTGTGGGCCACTTCAGCGAGGCCCCTGGCACCAGCGCCTTCGGCACCGCCGAGAGACCCGAGTTCGAGAGAATCC
TGAACGAGTGTAGGGCCGGCAGGCTGAACATGATCATCGTGTACGACGTGTCCCGGTTCAGCAGGCTGAAGGTGATGGACGCCATCCCTATCGTGTCCGA
GCTGCTGGCCCTGGGCGTGACCATCGTGTCCACCCAGGAAGGCGTCTTTAGACAGGGCAACGTGATGGACCTGATCCACCTGATCATGAGGCTGGACGCC
AGCCACAAGGAGAGCAGCCTGAAGAGCGCCAAGATCCTGGACACCAAGAACCTGCAGAGGGAGCTGGGCGGCTATGTGGGCGGCAAGGCCCCCTACGGCT
TCGAGCTGGTGTCCGAGACCAAGGAGATCACCCGGAACGGCAGGATGGTGAACGTGGTGATCAACAAGCTGGCCCACAGCACCACCCCCCTGACCGGCCC
CTTCGAGTTTGAGCCCGACGTGATCAGGTGGTGGTGGCGGGAGATCAAGACCCACAAGCACCTGCCTTTCAAGCCCGGCAGCCAGGCCGCCATCCACCCC
GGCAGCATCACCGGCCTGTGTAAGAGAATGGACGCCGACGCCGTGCCCACCAGAGGCGAGACCATCGGCAAGAAAACCGCCAGCAGCGCCTGGGACCCCG
CCACCGTGATGAGAATCCTGAGGGACCCTAGGATCGCCGGCTTCGCCGCCGAGGTGATCTACAAGAAGAAGCCCGACGGCACCCCCACCACCAAGATCGA
GGGCTACAGAATCCAGAGAGACCCCATCACCCTGAGACCTGTGGAGCTGGACTGTGGCCCTATCATCGAGCCTGCCGAGTGGTACGAGCTGCAGGCCTGG
CTGGACGGCAGAGGCAGAGGCAAGGGCCTGAGCAGAGGCCAGGCCATCCTGAGCGCCATGGACAAGCTGTACTGTGAGTGTGGCGCCGTGATGACCAGCA
AGAGAGGCGAGGAGAGCATCAAGGACAGCTACCGGTGCCGGAGAAGAAAGGTGGTGGACCCCAGCGCCCCTGGCCAGCACGAGGGCACCTGTAATGTGAG
CATGGCCGCCCTGGACAAGTTCGTGGCCGAGCGGATCTTCAACAAGATCCGGCACGCCGAGGGCGACGAGGAGACCCTGGCCCTGCTGTGGGAGGCCGCC
AGAAGATTCGGCAAGCTGACCGAGGCCCCCGAGAAGAGCGGCGAGAGGGCCAACCTGGTGGCCGAGAGAGCCGACGCCCTGAACGCCCTGGAGGAGCTGT
ACGAGGACAGAGCCGCCGGAGCCTATGACGGCCCTGTGGGCAGGAAGCACTTCAGAAAGCAGCAGGCCGCCCTGACCCTGAGACAGCAGGGCGCCGAGGA
AAGACTGGCCGAGCTGGAGGCCGCCGAGGCCCCTAAGCTGCCCCTGGATCAGTGGTTCCCCGAGGATGCCGACGCCGACCCCACCGGCCCCAAGTCCTGG
TGGGGCAGAGCCAGCGTGGACGACAAGAGGGTGTTCGTGGGCCTGTTCGTGGATAAGATCGTGGTGACCAAGAGCACCACCGGCAGGGGCCAGGGCACCC
CCATCGAGAAGAGAGCCAGCATCACCTGGGCCAAGCCTCCCACCGACGACGACGAGGATGACGCCCAGGACGGCACCGAGGACGTGGCCGCCCCTAAGAA
AAAGCGGAAAGTGAGATCGGAGGGCAGGGGAAGTCTTCTAACATGCGGGGACGTGGAGGAAAATCCCGGGCCCAGATCTACCCAGCTTTCTTGTACAAAG


TTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGCCATCCAGCTGATATCCCCTA
TAGTGAGTCGTATTACATGGTCATAGCTGTTTCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCA
TGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGG
ATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTT
TCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCAT
TTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATA
TTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCA
ATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTG
CCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAG
TCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTAT
TGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACG
CTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA
CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGG
ATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATG
AGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAAC
GCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT ••• 4455

## pROSA26 Dest

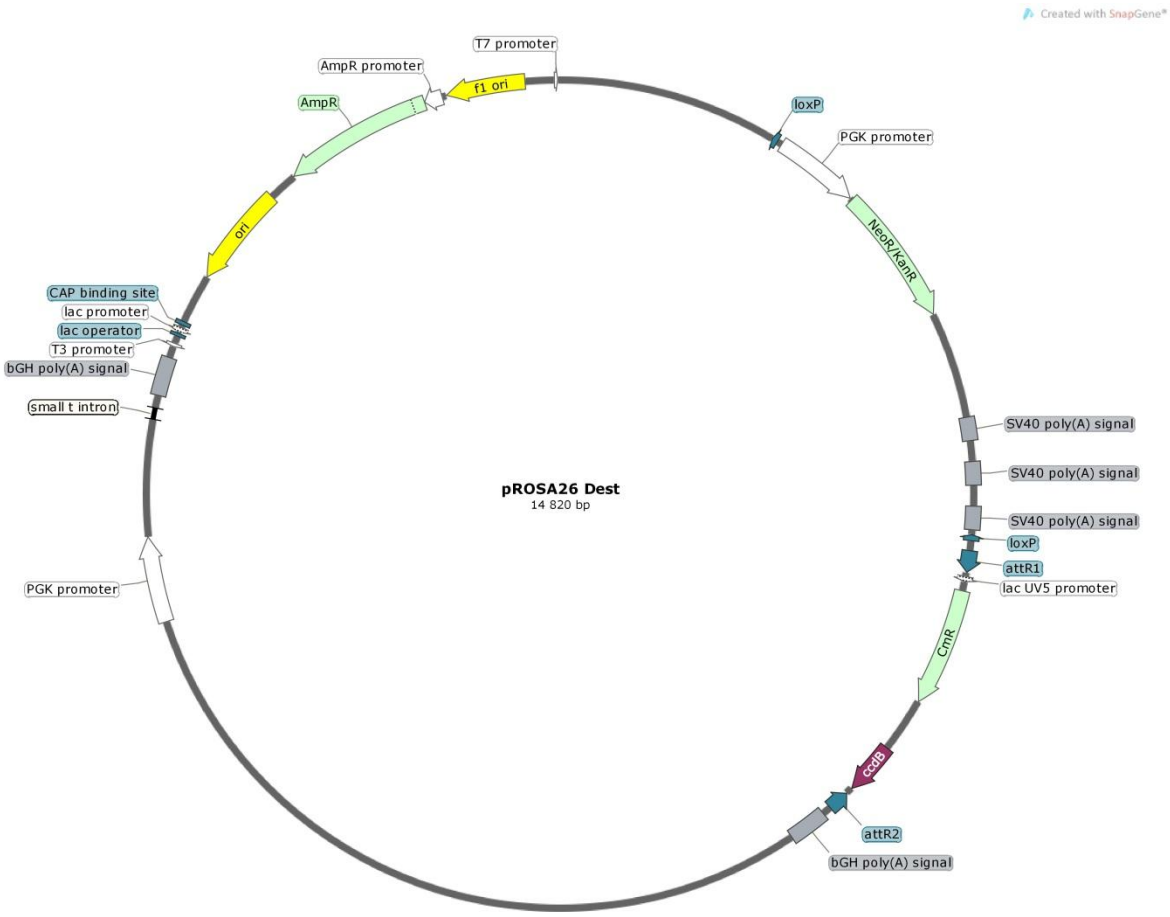
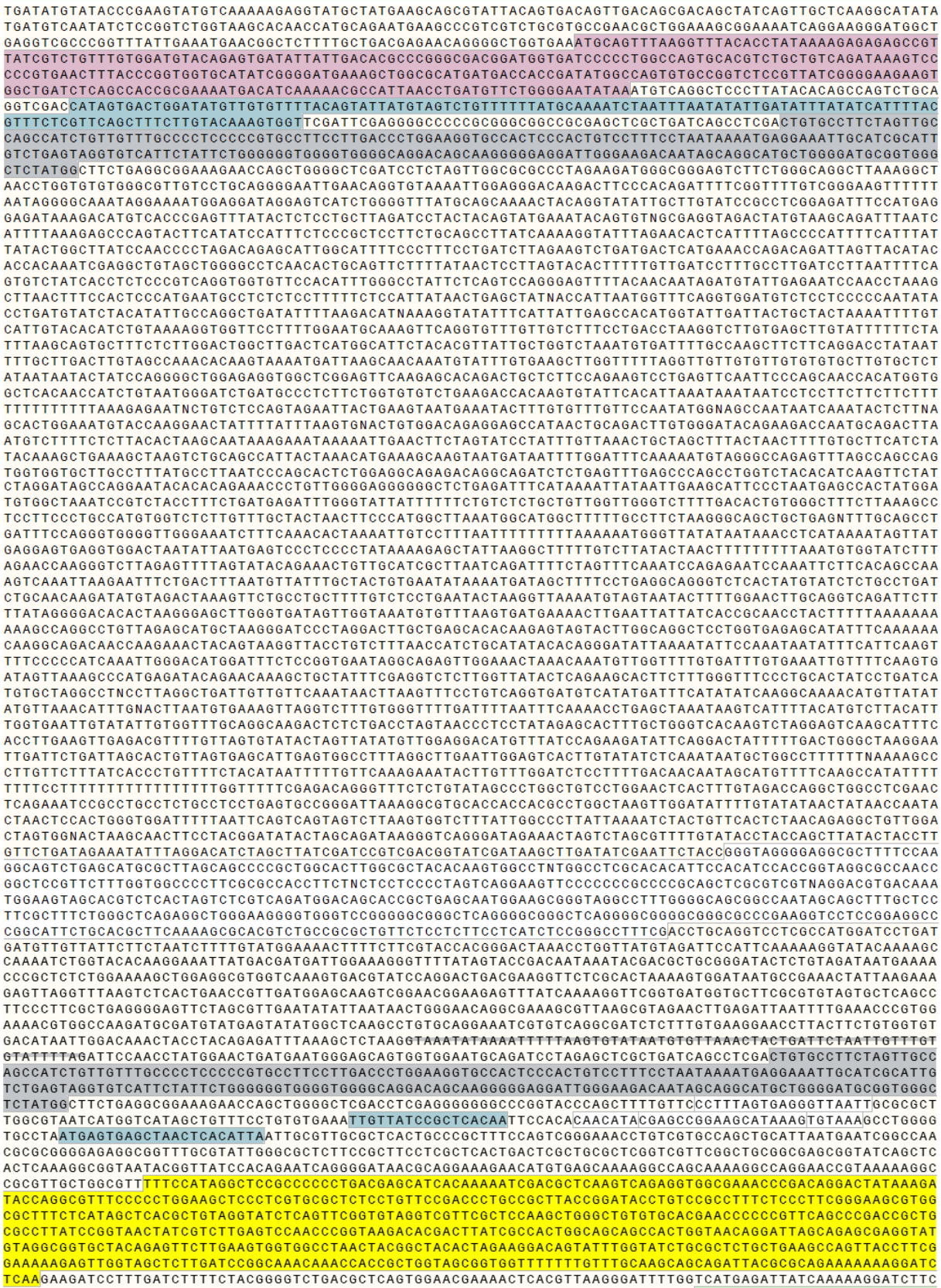
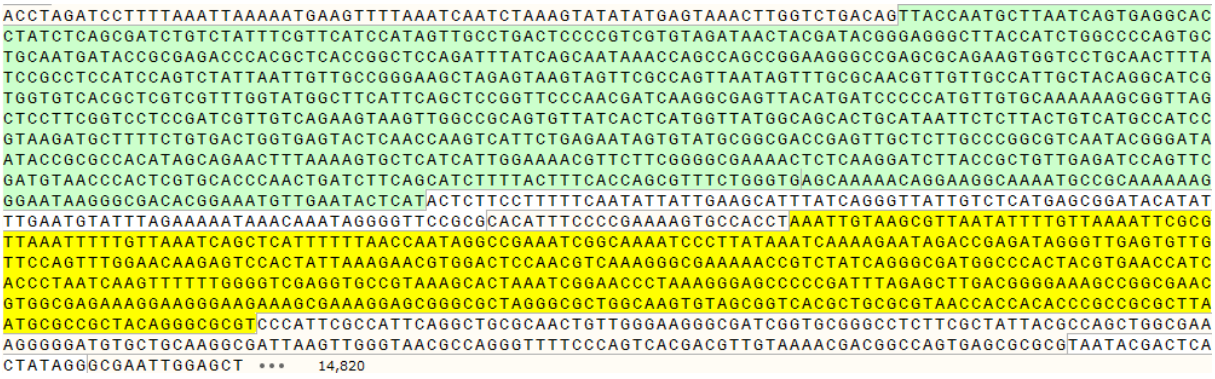



**Figure A16.** pROSA26 Dest vector map. Features were automatically annotated using SnapGene® Software.

CCCCGCGGCAGGCCCTCCGAGCGTGGTGGAGCCGTTCTGTGAGACAGCCGGGTACGAGTCGTGACGCTGGAAGGGGCAAGCGGGTGGTGGGCAGGAATGC
GGTCCGCCCTGCAGCAACCGGAGGGGGAGGGAGAAGGGAGCGGAAAAGTCTCCACCGGACGCGGCCATGGCTCGGGGGGGGGGGGGCAGCGGAGGANCGC
TTCCGGCCGACGTCTCGTCGCTGATTGGCTTNTTTTCCTCCCGCCGTGTGTGAAAACACAAATGGCGTGTTTTGGTTGGCGTAAGGCGCCTGTCAGTTAA
CGGCAGCCGGAGTGCGCAGCCGCCGGCAGCCTCGCTCTGCCCACTGGGTGGGGCGGGAGGTAGGTGGGGTGAGGCGAGCTGNACGTGCGGGCGCGGTCGG
CCTCTGGCGGGGCGGGGGAGGGGAGGGAGGGTCAGCGAAAGTAGCTCGCGCGCGAGCGGCCGCCCACCCTCCCCTTCCTCTGGGGGAGTCGTTTTACCCG
CCGCCGGCCGGGCCTCGTCGTCTGATTGGCTCTCGGGGCCCAGAAAACTGGCCCTTGCCATTGGCTCGTGTTCGTGCAAGTTGAGTCCATCCGCCGGCCA
GCGGGGGCGGCGAGGAGGCGCTCCCAGGTTCCGGCCCTCCCCTCGGCCCCGCGCCGCAGAGTCTGGCCGCGCGCCCCTGCGCAACGTGGCAGGAAGCGCG
CGCTGGGGGCGGGGACGGGCAGTAGGGCTGAGCGGCTGCGGGGCGGGTGCAAGCACGTTTCCGACTTGAGTTGCCTCAAGAGGGGCGTGCTGAGCCAGAC
CTCCATCGCGCACTCCGGGGAGTGGAGGGAAGGAGCGAGGGCTCAGTTGGGCTGTTTTGGAGGCAGGAAGCACTTGCTCTCCCAAAGTCGCTCTGAGTTG
TTATCAGTAAGGGAGCTGCAGTGGAGTAGGCGGGGAGAAGGCCGCACCCTTCTCCGGAGGGGGGAGGGGAGTGTTGCAATACCTTTCTGGGAGTTCTCTG
CTGCCTCCTGGCTTCTGAGGACCGCCCTGGGCCTGGGAGAATCCCTTGCCCCCTCTTCCCCTCGTGATCTGCAACTCCAGTCTTTCTAGTTAATTAAGGG
ATCTGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACTTATCCTGTCCCTTTTTTTTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGT
CTTTCCAGTGGGGATCGACGGTATCGTAGAGTCGAGGCCGCTCTAGAACTAGTGGATCCGGAACCCTTAATATAACTTCGTATAATGTATGCTATACGAA
GTTATTAGGTCCCTCGACCTGCAGGAATTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGCACTTG
GCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTC
CTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGATGGACAG
CACCGCTGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGTCC
GGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGCGGGCGCGAAGGTCCTCCCGAGGCCCGGCATTCTCGCACGCTTCAAAAGCGCACGTCTGCCGCGCT
GTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCAGCCAATATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGT
GGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTC
AAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACG
TTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCAT
GGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAA
GCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGATG
ATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGA
CCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCC
GATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGGATCCGCTGTAAGTCTGCAGAAATTGATGATCTATTAAACAATAAAGAT




ACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCAC
CTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
TGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTA
TCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCG
TGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAG
CTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCC
GTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATA
ATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTC
GATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAG
GGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATAT
TTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCG
TTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTG
TTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATC
ACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAAC
GTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTA
ATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAA
AGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCA
CTATAGGGCGAATTGGAGCT ••• 14,820

## pROSA26-DV2

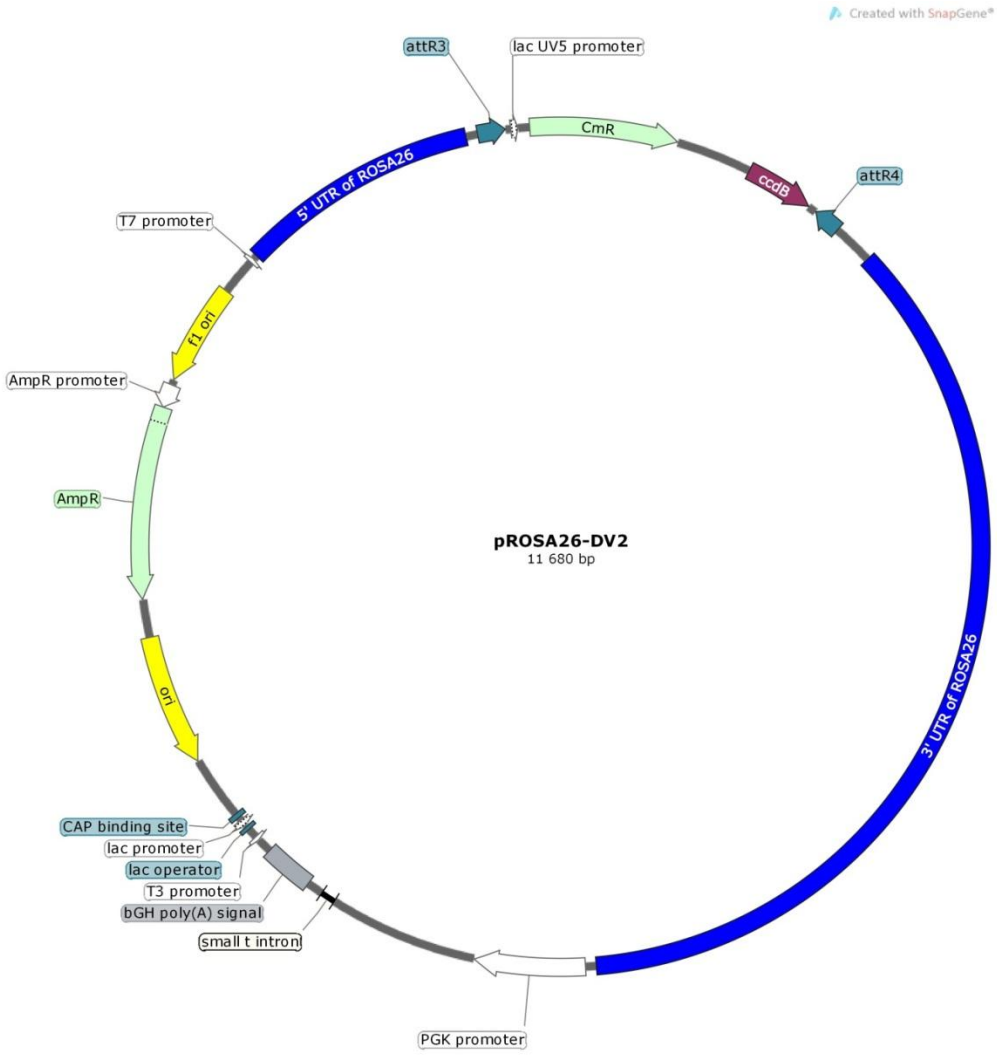
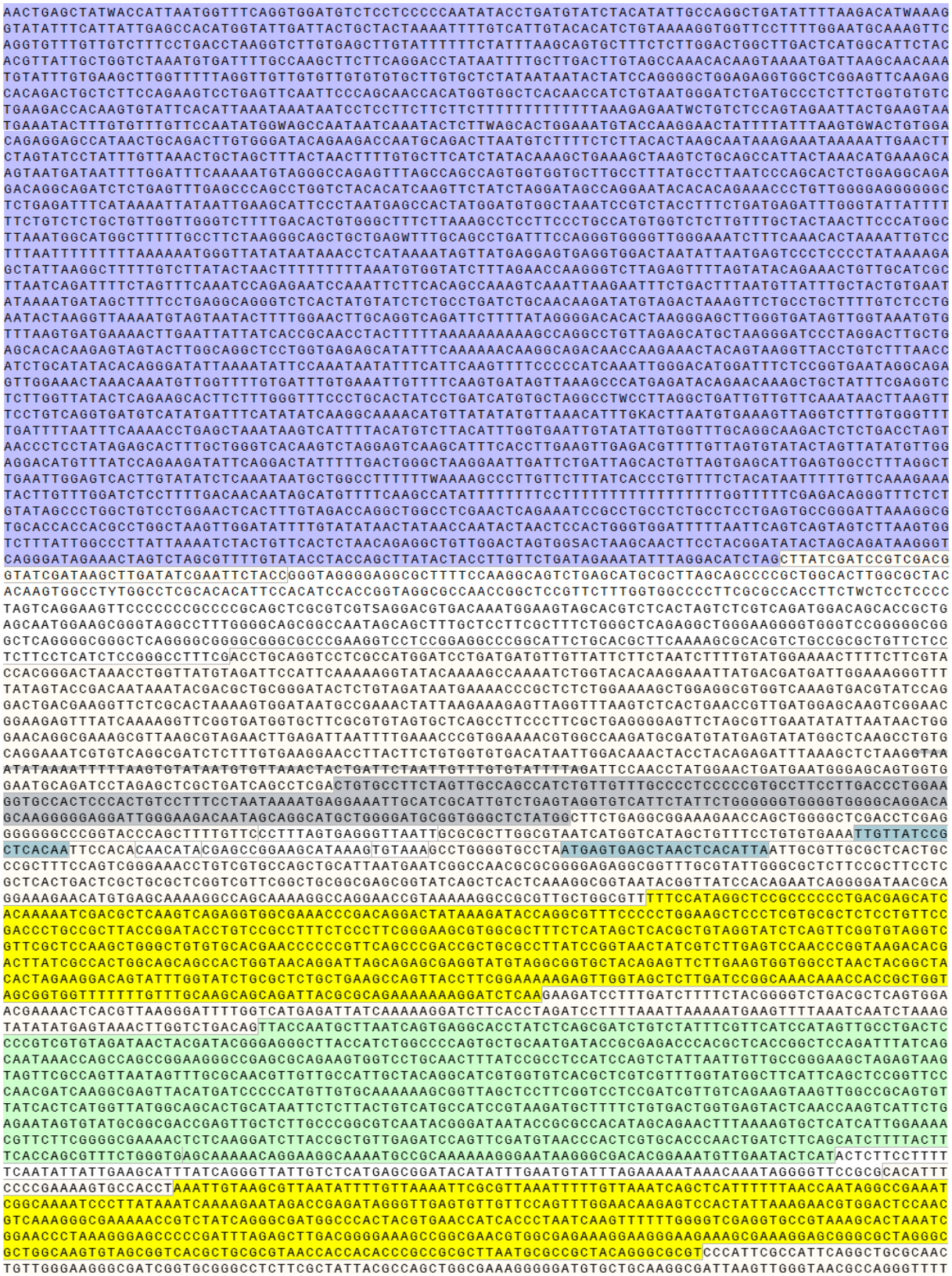
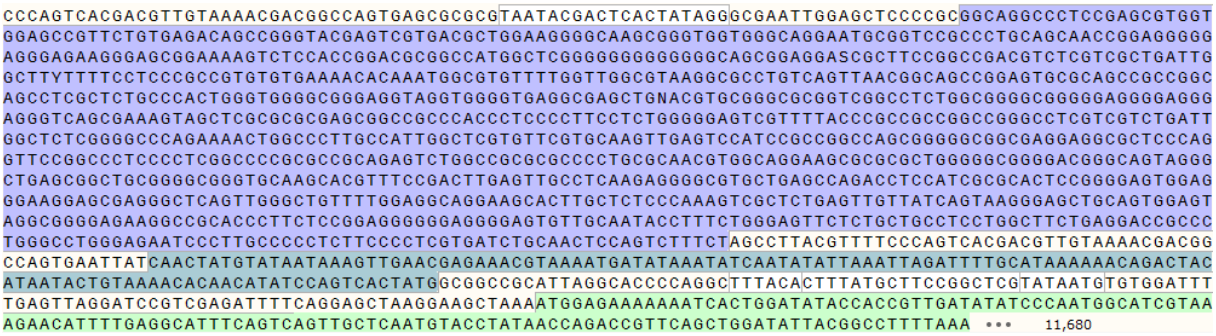



**Figure A17.** pROSA26-DV2 vector map. Features were automatically annotated using SnapGene® Software.

AGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGA
CGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGAT
TTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCT
CAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCA
AGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGC
GATGAGTGGCAGGGCGGGGCGTAATCTAGAGGATCCGGCTTACTAAAAGCCAGATAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATAT
ATACTGATATGTATACCCGAAGTATGTCAAAAAGAGGTATGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCGACAGCTATCAGTTGCTCAAGGCA
TATATGATGTCAATATCTCCGGTTCGGTAAGCACAACCATGCAGAATGAAGCCCGTCGTCTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAAGGGAT
GGCTGAGGTCGCCCCGGTTTATTGAAATGAACGGCTCTTTTGCCGACGAGAACAGGGGCTGGTGAAATGCAGTTTAAGGTTTACACCTATAAAAGAGAGAG
CCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAA
GTCCCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAG
AAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGGCTCCCTTATACACAGCCAGTC
TGCAGGTCGACCATAGTGACTGGATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCATT
TTACGTTTCTCGTTCAACTTTTCTATACAAAGTTGATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCGTAATTTAA
ATGGCGCGCCGGCTAGAAGATGGGCGGGAGTCTTCTGGGCAGGCTTAAAGGCTAACCTGGTGTGTGGGCGTTGTCCTGCAGGGGAATTGAACAGGTGTAA
AATTGGAGGGACAAGACTTCCCACAGATTTTCGGTTTTGTCGGGAAGTTTTTTAATAGGGGCAAATAGGAAAATGGAGGATAGGAGTCATCTGGGGTTTA
TGCAGCAAAACTACAGGTATATTGCTTGTATCCGCCTCGGAGATTTCCATGAGGAGATAAAGACATGTCACCCGAGTTTATACTCTCCTGCTTAGATCCT
ACTACAGTATGAAATACAGTGTYGCGAGGTAGACTATGTAAGCAGATTTAATCATTTTAAAGAGCCCAGTACTTCATATCCATTTCTCCCGCTCCTTCTG
CAGCCTTATCAAAAGGTATTTAGAACACTCATTTTAGCCCCATTTTCATTTATTATACTGGCTTATCCAACCCCTAGACAGAGCATTGGCATTTTCCCTT
TCCTGATCTTAGAAGTCTGATGACTCATGAAACCAGACAGATTAGTTACATACACCACAAATCGAGGCTGTAGCTGGGGCCTCAACACTGCAGTTCTTTT
ATAACTCCTTAGTACACTTTTTGTTGATCCTTTGCCTTGATCCTTAATTTTCAGTGTCTATCACCTCTCCCGTCAGGTGGTGTTCCACATTTGGGCCTAT
TCTCAGTCCAGGGAGTTTTACAACAATAGATGTATTGAGAATCCAACCTAAAGCTTAACTTTCCACTCCCATGAATGCCTCTCTCCTTTTTCTCCATTAT




CCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCCCGCGGCAGGCCCTCCGAGCGTGGT
GGAGCCGTTCTGTGAGACAGCCGGGTACGAGTCGTGACGCTGGAAGGGGCAAGCGGGTGGTGGGCAGGAATGCGGTCCGCCCTGCAGCAACCGGAGGGGG
AGGGAGAAGGGAGCGGAAAAGTCTCCACCGGACGCGGCCATGGCTCGGGGGGGGGGGGGCAGCGGAGGASCGCTTCCGGCCGACGTCTCGTCGCTGATTG
GCTTYTTTTCCTCCCGCCGTGTGTGAAAACACAAATGGCGTGTTTTGGTTGGCGTAAGGCGCCTGTCAGTTAACGGCAGCCGGAGTGCGCAGCCGCCGGC
AGCCTCGCTCTGCCCACTGGGTGGGGCGGGAGGTAGGTGGGGTGAGGCGAGCTGNACGTGCGGGCGCGGTCGGCCTCTGGCGGGGCGGGGGAGGGGAGGG
AGGGTCAGCGAAAGTAGCTCGCGCGCGAGCGGCCGCCCACCCTCCCCTTCCTCTGGGGGAGTCGTTTTACCCGCCGCCGGCCGGGCCTCGTCGTCTGATT
GGCTCTCGGGGCCCAGAAAACTGGCCCTTGCCATTGGCTCGTGTTCGTGCAAGTTGAGTCCATCCGCCGGCCAGCGGGGGCGGCGAGGAGGCGCTCCCAG
GTTCCGGCCCTCCCCTCGGCCCCGCGCCGCAGAGTCTGGCCGCGCGCCCCTGCGCAACGTGGCAGGAAGCGCGCGCTGGGGGCGGGGACGGGCAGTAGGG
CTGAGCGGCTGCGGGGCGGGTGCAAGCACGTTTCCGACTTGAGTTGCCTCAAGAGGGGCGTGCTGAGCCAGACCTCCATCGCGCACTCCGGGGAGTGGAG
GGAAGGAGCGAGGGCTCAGTTGGGCTGTTTTGGAGGCAGGAAGCACTTGCTCTCCCAAAGTCGCTCTGAGTTGTTATCAGTAAGGGAGCTGCAGTGGAGT
AGGCGGGGAGAAGGCCGCACCCTTCTCCGGAGGGGGGAGGGGAGTGTTGCAATACCTTTCTGGGAGTTCTCTGCTGCCTCCTGGCTTCTGAGGACCGCCC
TGGGCCTGGGAGAATCCCTTGCCCCCTCTTCCCCTCGTGATCTGCAACTCCAGTCTTTCTAGCCTTACGTTTTCCCAGTCACGACGTTGTAAAACGACGG
CCAGTGAATTATCAACTATGTATAATAAAGTTGAACGAGAAACGTAAAATGATATAAATATCAATATATTAAATTAGATTTTGCATAAAAAACAGACTAC
ATAATACTGTAAAACACAACATATCCAGTCACTATGGCGGCCGCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGATTT
TGAGTTAGGATCCGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAA
AGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTAAA ••• 11,680

## pROSA26-DV2-CAGGS-PHIC31-NEOPA

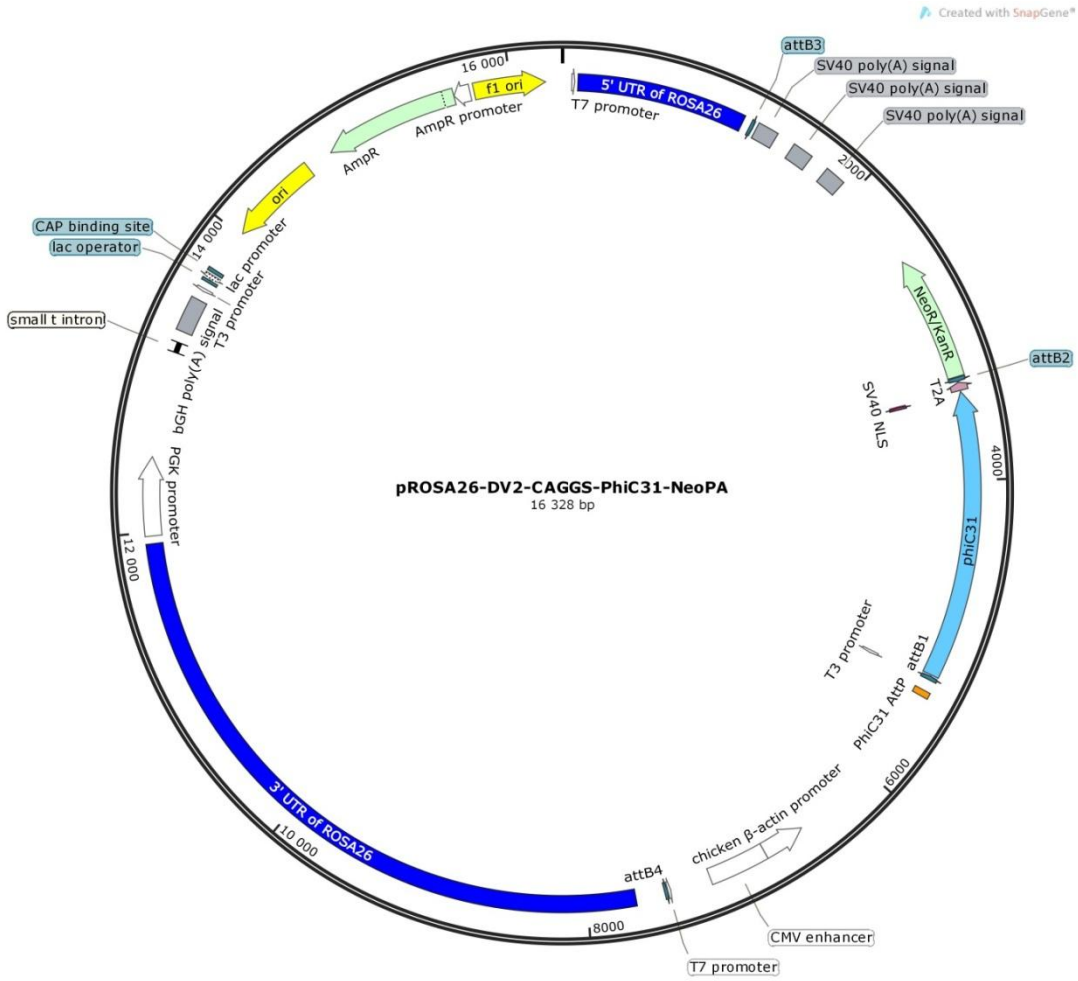
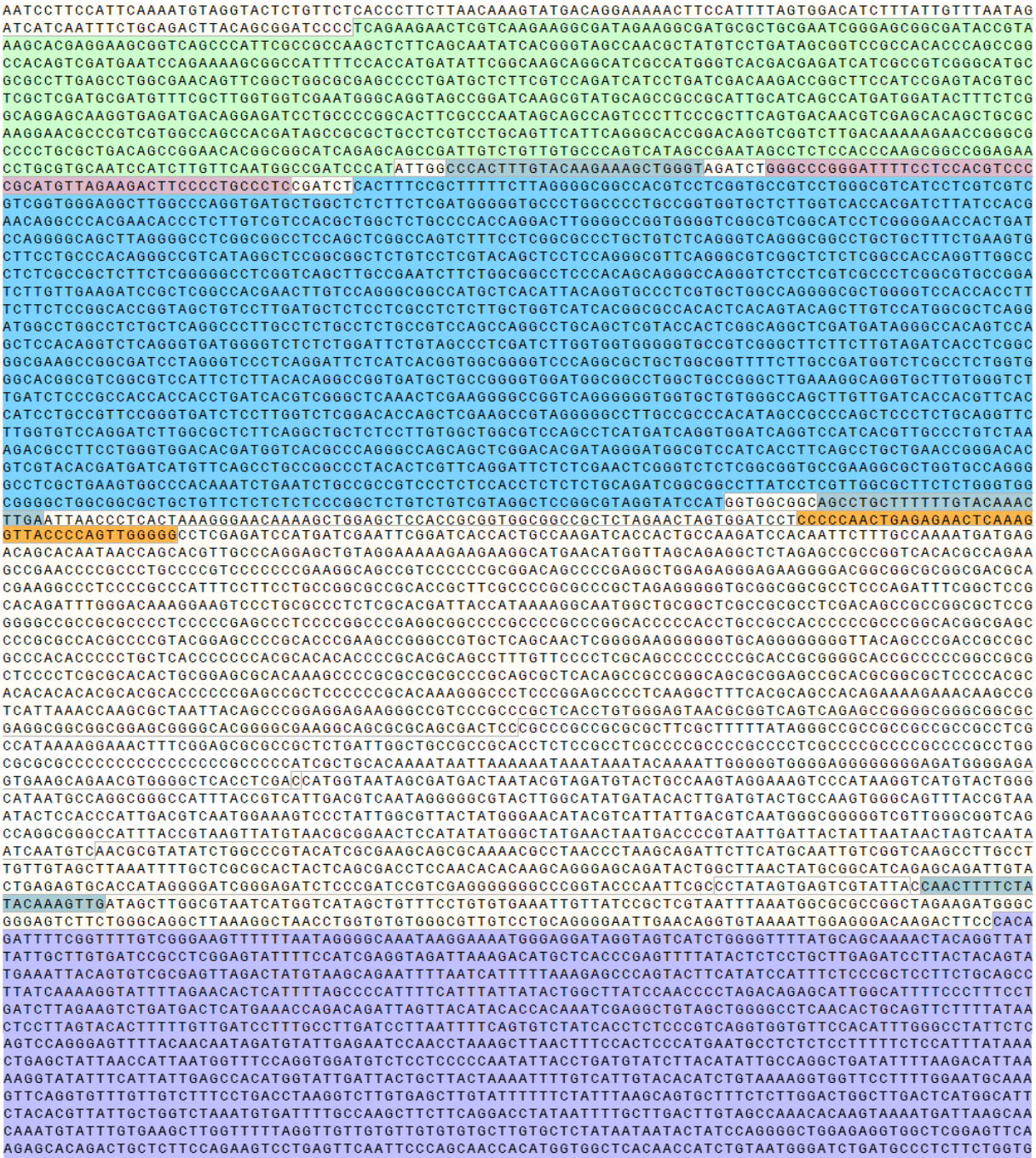
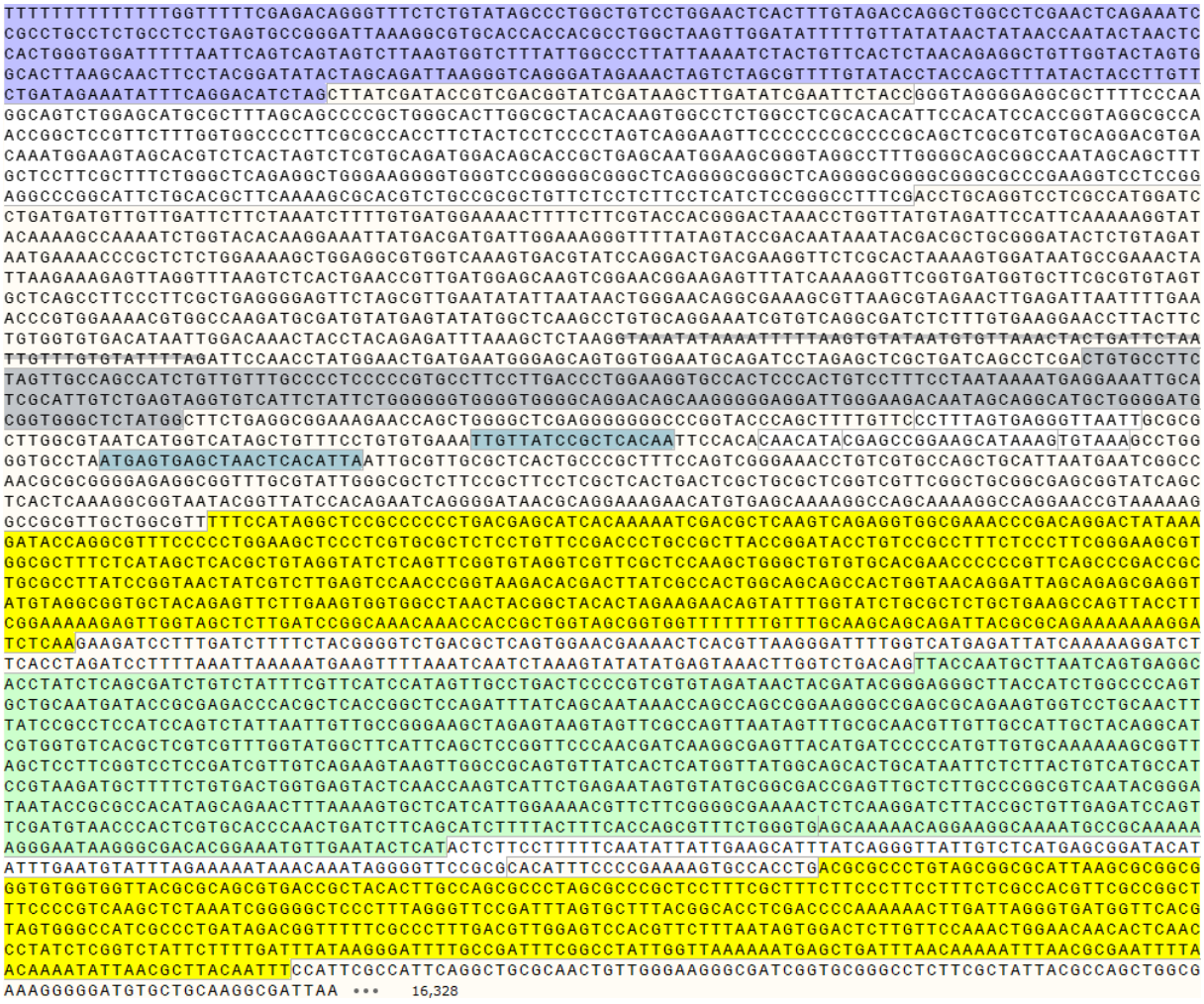



**Figure A18.** pRosa26-DV2-CAGGS-PhiC31-NeoPA vector map. Features were automatically annotated using SnapGene® Software.

GTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCG
CGGCAGGCCCTCCGAGCGTGGTGGAGCCGTTCTGTGAGACAGCCGGGTACGAGTCGTGACGCTGGAAGGGGCAAGCGGGTGGTGGGCAGGAATGCGGTCC
GCCCTGCAGCAACCGGAGGGGGAGGGAGAAGGGAGCGGAAAAGTCTCCACCGGACGCGGCCATGGCTCGGGGGGGGGGGCAGCGGAGGAGCGCTTCCGGC
CGACGTCTCGTCGCTGATTGGCTTCTTTTCCTCCCGCCGTGTGTGAAAACACAAATGGCGTGTTTTGGTTGGCGTAAGGCGCCTGTCAGTTAACGGCAGC
CGGAGTGCGCAGCCGCCGGCAGCCTCGCTCTGCCCACTGGGTGGGGCGGGAGGTAGGTGGGGTGAGGCGAGCTGNACGTGCGGGCGCGGTCGGCCTCTGG
CGGGGCGGGGGAGGGGAGGGAGGGTCAGCGAAAGTAGCTCGCGCGCGAGCGGCCGCCCACCCTCCCCTTCCTCTGGGGGAGTCGTTTTACCCGCCGCCGG
CCGGGCCTCGTCGTCTGATTGGCTCTCGGGGCCCAGAAAACTGGCCCTTGCCATTGGCTCGTGTTCGTGCAAGTTGAGTCCATCCGCCGGCCAGCGGGGG
CGGCGAGGAGGCGCTCCCAGGTTCCGGCCCTCCCCTCGGCCCCGCGCCGCAGAGTCTGGCCGCGCGCCCCTGCGCAACGTGGCAGGAAGCGCGCGCTGGG
GGCGGGGACGGGCAGTAGGGCTGAGCGGCTGCGGGGCGGGTGCAAGCACGTTTCCGACTTGAGTTGCCTCAAGAGGGGCGTGCTGAGCCAGACCTCCATC
GCGCACTCCGGGGAGTGGAGGGAAGGAGCGAGGGCTCAGTTGGGCTGTTTTGGAGGCAGGAAGCACTTGCTCTCCCAAAGTCGCTCTGAGTTGTTATCAG
TAAGGGAGCTGCAGTGGAGTAGGCGGGGAGAAGGCCGCACCCTTCTCCGGAGGGGGGAGGGGAGTGTTGCAATACCTTTCTGGGAGTTCTCTGCTGCCTC
CTGGCTTCTGAGGACCGCCCTGGGCCTGGGAGAATCCCTTCCCCCTCTTCCCTCGTGATCTGCAACTCCAGTCTTTCTAGCCTTACGTTTTCCCAGTCAC
GACGTTGTAAAACGACGGCCAGTGAATTATCAACTATGTATAATAAAGTTGGGTTCCGGATCAGCTTGATGGGGATCCAGACATGATAAGATACATTGAT
GAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATA
AACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAT
GGCTGATTATGATCCTCTAGAGTCGCAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTA
TTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCA
GGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGCAGATCCAGACATGATAAGAT
ACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAG
CTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAA
TGTGGTATGGCTGATTATGATCCTCTAGAGTCGCAGATCCTCTAGAGTCGCAGATCTGCAAGCTTCTGATGGAATTAGAACTTGGCAAAACAATACTGAG
AATGAAGTGTATGTGGAACAGAGGCTGCTGATCTCGTTCTTCAGGCTATGAAACTGACACATTTGGAAACCACAGTACTTAGAACCACAAAGTGGGAATC
AAGAGAAAAACAATGATCCCACGAGAGATCTATAGATCTATAGATCATGAGTGGGAGGAATGAGCTGGCCCTTAATTTGGTTTTGCTTGTTTAAATTATG
ATATCCAACTATGAAACATTATCATAAAGCAATAGTAAAGAGCCTTCAGTAAAGAGCAGGCATTTATCTAATCCCACCCCACCCCCACCCCCGTAGCTCC




16,328

## pUC57-Amp_VEcad_GTSmCherry-2A-H2BeGFPpA_FRT.Neo.FRTpA

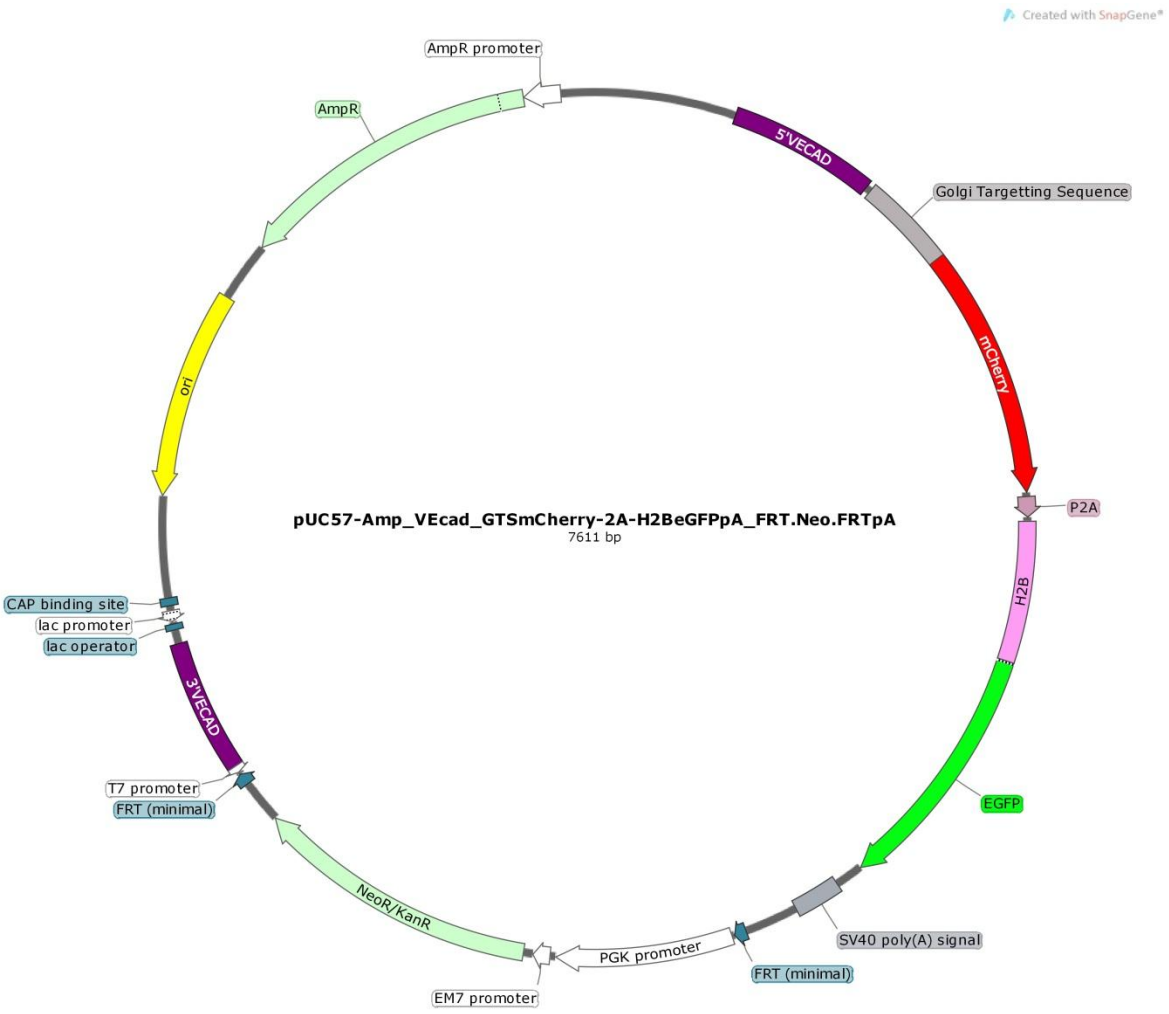



**Figure A19.** pUC57-Amp_VEcad_GTSmCherry-2A-H2BeGFPpA_FRT.Neo.FRTpA vector map. Features were automatically annotated using SnapGene® Software.

```
TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCG
TCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATA
CCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTAT
TACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATT
CAGCACTCTCTGGAGGTGACTGGGTGTGGTCAGAGGCTGGGGCTTCATTCTCAGCTGAAGTTAACCTCTCTGTGCCTCAGTACCCTCTTTTATAAAACAG
TGATATTCTGAGTGCTCCCATGATGCTTTGGGAGCTTGGACATCAGTATTCCAGGTATCCTCCCCGGGTTCCTGGGGATCTCAGACCCCAACACCTGCTA
CCATATCATCCTACGTGGGTATAAGGAGAAATCGTTTCTGGGCTTGCCTTCCCTACTTAGACTAGCTTACCATTCTGCTGGTGTCAGCTGAGTGGCAGAT
AAAGAAGGAGAGAGACAAAGAAACTTTGGCTCCTAGAGTGAGCCACTCGTTCCATAGGACAGTGGGCACTGGGTCCTGATGGTGCCTATCCTCTTTCCCC
AGATCAGCTCCTCCACGAAGGGCTAGCGCCGCCAGCATGAGGCTTCGGGAGCCGCTCCTGAGCGGCAGCGCCGCGATGCCAGGCGCGTCCCTACAGCGGG
CCTGCCGCCTGCTCGTGGCCGTCTGCGCTCTGCACCTTGGCGTCACCCTCGTTTACTACCTGGCTGGCCGCGACCTGAGCCGCCTGCCCCAACTGGTCGG
AGTCTCCACACCGCTGCAGGGCGGCTCGAACAGTGCCGCCGCCATCGGGCAGTCCTCCGGGGAGCTCCGGACCGGAGGGGCCAAGGATCCACCGGTCGCC
ACCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCG
AGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCT
GTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAG
CGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCA
ACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGAT
CAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC
GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACG
AGCTGTACAAGGTCGACGATCCAGCCACGAACTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGAAAACCCCGGTCCTCGCAGAAGATCTATGCCAGA
GCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAG
AGCTATTCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCATGGGCATCATGAATTCGTTTGTGAACGACA
TTTTCGAGCGCATCGCAGGTGAGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGAGATCCAGACGGCCGTGCGCCTGCTGCT
GCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAGGATCCACCGGTCGCCACCATGGTGAGCAAG
```


7611